THE INVERSE OPTIMAL LINEAR REGULATOR PROBLEM

By

Joseph Marcus Elder, Jr.

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA
1972

To my grandfather
Roy W. Estridge
who never doubted
and to my wife Dina
who never despaired

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to the members of his supervisory committee: Dr. T. E. Bullock, chairman; Dr. W. H. Boykin, and Dr. A. E. Durling for their counsel and patience. Special thanks are due to Dr. Bullock who gave unselfishly of his time and energy; and abject apologies are due to his wife, Jeannette, who endured the many late hours and missed dinners.

Thanks also are due to Dr. C. V. Shaffer who had an uncanny knack for dispatching the multitude of small problems encountered along the way, to Dr. L. E. Jones who was always ready to help, and to Mr. Y. T. Parker for a great deal of food for thought.

The author is grateful to the University of Florida Graduate School, the Department of Electrical Engineering, and the National Aeronautics and Space Administration for financial assistance.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

Abstract of Dissertation Presented to the
Graduate Council of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of Doctor of Philosophy

THE INVERSE OPTIMAL LINEAR REGULATOR PROBLEM

By

Joseph Marcus Elder, Jr.

March, 1972

Chairman: Dr. T. E. Bullock
Major Department· Electrical Engineering

The principal objective of this research is to extend the results of R. E. Kalman's original analysis of the inverse optimal control problem in order to create a design tool which allows the designer to employ the powerful and sophisticated techniques of optimal control theory when the correct choice of a performance index is not apparent. Specifically, methods are developed for the design of optimal linear regulators (both continuous and discrete) which satisfy classical performance specifications in addition to the minimality of a quadratic functional. This permits the great computing power of optimal control theory to be brought to bear on problems which previously could be accommodated only by the cut-and-try methods of classical design schemes. In passing, insights into the basic nature of optimal regulators are developed in terms of classical concepts, unifying some important notions in classical and modern control theory.

The contributions of this research can be summarized as follows:

1. A complete theoretical investigation of the inverse optimal linear regulator problem for quadratic performance indices with positive semidefinite state

weighting matrices and scalar input systems is reviewed and preliminary results for the case where the weighting matrix is allowed to be sign indefinite are given.

2. The equivalence of performance indices for scalar input linear quadratic loss problems is resolved and a procedure for generating the entire equivalence class of cost functions equivalent to a given performance index is developed.

3. Practical numerical methods for determination of system optimality and computation of solutions to the scalar input inverse problem are discussed.

4. Some of the effects of specific elements of the performance index on optimal system performance and pole locations are determined.

5. The problem of designing optimal systems to meet classical performance specifications is encountered and some definitive results obtained.

6. A solution to the problem of designing a sampled-data controller to approximate the performance of continuous controller is given.

CHAPTER I

INTRODUCTION

1.1 Background

In 1964 R. E. Kalman published a paper [K1][1] which has come to have considerable impact on the theory of optimal control. It dealt not with the conventional optimal control problem of computing a system trajectory which extremizes a specified performance index, but rather with the "inverse" problem of determining what performance indices, if any, are extremized by a specified control. Kalman restricted his attention to the case where the system is linear and the integral performance index is quadratic in the states and control. No immediate direct application was made of Kalman's results, but they were of great value in the application and interpretation of his earlier analysis [K2] of what is now called the "linear quadratic loss" optimization problem.

Quadratic functionals have long been studied in the calculus of variations [G1] but not until Kalman was the minimization of a functional quadratic in the states and input of a linear dynamical system considered in the context of an optimal control problem. Other authors, notably S. S. L. Chang [C1] and Newton, Gould, and Kaiser [N1],

---

[1] Brackets contain reference information. A letter followed by a numeral indicates, respectively, the first letter of the first author's surname and the order of appearance within the given alphabetic group of the specific reference.

had since the middle 1950's been occupied with the "analytical design" of linear control systems through the use of a performance index quadratic in the error (i.e., the difference between actual and desired system responses to a specified input). Employment of Parseval's Theorem and spectral factorization [N1] resulted in a solution strongly related to the optimal linear filter of Wiener [W1] of almost fifteen years earlier. The solutions for the ISE (integral of the square of the error) problems were very cumbersome to compute and seldom applied to systems of greater than third order.

In the early 1960's the fledgling field of optimal control theory underwent a metamorphosis. Difficult aerospace problems had arisen which could only be conveniently approached as optimal control problems. Rapidly there began appearing almost as many different optimal control formulations as there were proponents. Superficially, optimal control theory appeared to be a panacea; however, formidable problems in computing and often even greater difficulties in the implementation of some optimal designs limited their utility. In general, the computation of optimal controls requires the use of iterative algorithms which converge, at best, slowly. Further, optimal controls are usually open-loop in nature.

The linear quadratic loss problem suffers from neither of these difficulties: the solution is numerically straightforward and always results in a linear feedback control law which is easily implemented; in addition both continuous and sampled-data controllers can be accommodated. If the system is to operate in a noisy environment, an extension of the quadratic loss formulation, the Kalman-Bucy filter, can easily be included in the design.

Unfortunately, performance indices as a whole seldom relate well to system design requirements which are given as classical time and frequency domain specifications. In this regard, the quadratic loss problem suffers as well.

The principal objective of this research is to extend those original results of Kalman in order to create a design tool which allows the designer to employ the powerful and sophisticated techniques of optimal control theory when the correct choice of a performance index is not apparent. Specifically, methods are developed for the design of optimal linear regulators (both continuous and discrete) which satisfy classical performance specifications in addition to the minimality of a quadratic functional. This would permit the great computing power of optimal control theory to be brought to bear on problems which previously could be accommodated only by the cut-and-try methods of classical design schemes. In passing, insights into the basic nature of optimal regulators are developed in terms of classical concepts, unifying some important notions in classical and modern control theory.

The original impetus for this investigation was the requirement to develop a method for designing a digital controller to replace an existing continuous compensator without significantly affecting system performance. It was believed that determination of a performance index minimized by the continuous system would allow for performance invariant design by computing the optimal control law for the sampled-data versions of the continuous system and performance index. In Chapter IV this scheme is discussed as a solution to a problem which will

undoubtedly occur with increasing frequency as digital controllers become more commonplace. It was in the course of this investigation that the potential of the inverse problem for regulator design of wider latitude was discovered.

1.2 Survey of Previous Work

Perhaps the first attempts at relating optimal control theory and classical design were early (c. 1950) investigations of so-called "standard forms" (e.g., [G2]). The object was to tabulate forms of closed-loop system transfer functions which were optimal with respect to a specified performance measure (for instance, ISE) and input. This approach was rather restrictive and difficult to apply to problems of interest. Its impact was nonetheless considerable and a recent paper by Rugh [R1] indicates that the linear quadratic loss problem has much in common with these early results.

Shortly after the linear quadratic loss problem became widely known, the task of selecting performance indices which result in optimal systems with desired characteristics was attacked primarily on an experimental basis. The hope was that massive experience and some insight into the mechanics of computing the optimal control laws would lead to guidelines for the choice of a quadratic performance index. One such procedure, developed by Tyler and Tuteur [T1] consisted of computing root-locus plots as functions of weighting terms in the performance index and choosing a suitable compromise.

Another study of optimal systems in classical terms was made in 1965 by R. J. Leake [L1]. Leake, in developing a computational scheme for solutions to the linear quadratic loss problem, shows how

an estimate of optimal system bandwidth may be made without computing the optimal control and comments on an intriguing geometric interpretation (in the complex plane) of Kalman's criteria for optimality.

An important step in the analysis that follows will be the determination of when two different quadratic performance indices applied to the same plant will result in identical optimal control laws. Kalman's paper on the inverse problem, by its nature, encounters the problem theoretically but does not consider it in detail. A partial practical solution to the equivalence problem appeared simultaneously with Kalman's paper. Wonham and Johnson observed in their paper on the linear quadratic problem with a bounded control [W2] that an arbitrary weighting matrix in the performance index could be replaced by a diagonal one which will result in the same optimal control. Their result arises from the observation that when the system is expressed in a canonical form (companion matrix form [R2]) the performance index may be reduced by repeated integration by parts to a diagonal form; they fail to realize, however, that the diagonalized version may no longer possess a solution. Kalman and Englar later recognize [K3, p. 304-306] that, in the same canonical form as used by Wonham and Johnson, certain terms may be discarded and equivalence maintained.

Beyond these early results and an occasional rediscovery of them (e.g., Kreindler and Hedrick [K4]) the study of equivalent quadratic performance indices has remained dormant.

In summary:

1. A great deal of experimental work has been done in hopes of relating optimal design to classical criteria and providing intuition into a procedure for choosing performance indices. Very little basic theoretical research has been invested in this area.
2. The practical problems of determining the optimality of an actual system and the computation of a performance index have not been considered.
3. Some elementary equivalence relations have been developed, but the difficulties arising from naïvely applying them have been generally overlooked. Nor have the advantages of one equivalent form over another been explored.
4. No real effort has been made to use the study of the inverse problem as a foundation for linear regulator design.

This work begins with a review of the linear plant, quadratic cost variational problem. The conditions for the solution to exist are discussed along with techniques for computing the optimal control.

The inverse problem for a linear plant and quadratic performance index is next considered. The conditions for optimality are developed, the meaning of optimal control is studied in terms of classical criteria and the equivalence of performance indices is resolved.

The fourth chapter develops computational procedures for the determination of solutions to the inverse problem and goes on to consider practical techniques for the design of optimal regulators which meet classical performance specifications.

CHAPTER II

THE LINEAR, QUADRATIC COST OPTIMIZATION PROBLEM

2.1 Introduction

As a first step toward the establishment of viable design techniques based on the linear, quadratic cost optimal control problem, the optimization problem itself must be reviewed in some detail. The great power, latitude and computational elegance of this optimal control formulation account in large part for its popularity as an object of study. In the investigation that follows its limitations will also have considerable impact.

This chapter will first define what is meant by the linear plant, quadratic cost optimization problem. A method for the computation of solutions which also provides a necessary and sufficient test for existence will then be considered. The chapter closes with a specific review of the particular subproblem which will be of principal interest in the remainder of this work, the time-invariant optimal linear regulator.

2.2 The Linear System, Quadratic Cost Problem

Consider an n-dimensional linear system with state feedback described by the state equation

$$\frac{dx}{dt} = Fx + Gu \tag{1}$$

where $$u = -K^T(t)x$$

and a companion performance index

$$J = \frac{1}{2}\int_{t_o}^{t_f} (x^T Qx + u^T Ru)dt, \quad Q = Q^T \text{ and } R = R^T \tag{2}$$

where u is m-dimensional; F, G, Q, and R are real, constant matrices; and K is a real (possibly time-varying) matrix; all are of appropriate dimensions.

It will be shown later (Appendix A) that several other performance indices can be accommodated in the framework established for this particular one. The $\frac{1}{2}$ preceding the integral in (2) is for algebraic simplicity in the present discussion and has no effect on the result of the optimization process; it will occasionally be deleted in the sequel without comment.

Taken together there are a variety of problems inferred by (1) and (2). The most obvious is the optimal control problem [K2]. i.e., compute the control law K(t), if any, which minimizes the performance index (2). A second, somewhat more obscure problem, is the so-called "inverse problem" of optimal control theory [K1,E1]. The "inverse problem" seeks to determine what, if any, performance index is minimized by a specific feedback control law. In the following, both of these topics will play an important role in uncovering the nature of optimal systems.

2.3 The Optimal Control Problem

The Euler-Lagrange Equations

The necessary conditions for a control u to minimize the performance index (2) is that the system equations (1) and the celebrated Euler-Lagrange equations [G1] be simultaneously satisfied, i.e.,

$$\dot{\lambda} = -\frac{\partial H}{\partial x}, \qquad \lambda(t_f) = 0 \tag{3}$$

$$\frac{\partial H}{\partial u} = 0 \tag{4}$$

where H is the Hamiltonian

$$H = \frac{1}{2} x^T Q x + \frac{1}{2} u^T R u + \lambda^T (Fx + Gu).$$

If R is non-singular, equation (4) (sometimes called the "stationarity condition" [B3]) determines a candidate for the optimal control, i.e.,

$$u(t) = -R^{-1} G^T \lambda(t) . \tag{5}$$

The simultaneous solution of equations (1) and (3) with the substitution of (5) results in the two-point boundary value problem:

$$\frac{d}{dt}\begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} F & -GR^{-1}G^T \\ -Q & -F^T \end{bmatrix}\begin{bmatrix} x \\ \lambda \end{bmatrix} \qquad \begin{aligned} x(t_o) &= x_o \text{ (given)}, \\ \lambda(t_f) &= 0 . \end{aligned} \tag{6}$$

The Riccati Equation

The optimal control problem for a linear system subject to a quadratic performance index[1] has been studied in great detail over the past several years (e.g., [K2], [K3]) and several eloquent solutions for the necessary conditions have been offered [B1]. Most of these solutions differ only in the approach taken to solve the two-point boundary value problem (6); for the purpose of this chapter the so-called "sweep method" [G1] is most illustrative and will be presented by way of review.

Hypothesize a matrix $P(t,t_f)$ such that

$$\lambda(t) = P(t,t_f)x(t), \tag{7}$$

then $P(t,t_f)$ would in effect provide a boundary value which is "swept back" in time to the initial time and the initial value for λ would simply be

$$\lambda(t_o) = P(t_o,t_f)x_o,$$

hence, the separation of the boundary values would be resolved. By requiring that the Euler-Lagrange equations be satisfied, it may be shown that

$$\left(\frac{dP}{dt} + PF + F^TP - PGR^{-1}G^TP + Q\right)x = 0.$$

Since $x(t)$ is arbitrary, $P(t,t_f)$ must satisfy

$$\frac{dP}{dt} = -PF - F^TP + PGR^{-1}G^TP - Q,\ P(t_f,t_f) = 0 \tag{8}$$

which is a matrix version of the familiar Riccati equation [D1].

---

[1]This problem is often referred to as the "optimal linear regulator" problem in the literature; however, this designation will be reserved for the infinite final time case here.

At first it may seem rather surprising that the solution to a 2n order two-point boundary value problem (6) can be obtained from the solution of an nth order non-linear equation (8), but recalling that the scalar Riccati equation was solved in elementary differential equation theory with the aid of a 2nd order linear differential equation relates it to a familiar problem.

Finally, substitution of (7) into (5) leads to a candidate for the optimal control (a control which satisfies the necessary conditions) in the form of a feedback control law, i.e.,

$$u = -K^T(t)x, \text{ where } K(t) = P(t,t_f)GR^{-1} \tag{9}$$

and P is the solution to the Riccati equation (8).

Before proceeding any further, another necessary condition is immediately available which has been previously ignored in this analysis. From observation of the performance index (2), it is obvious that R must be positive semidefinite; otherwise, a control could be hypothesized with sufficient high frequency content to have negligible effect on the states of the system, hence allowing the integral (2) to become unbounded (negatively). The requirement in the Euler-Lagrange equations that R be non-singular further constrains R to be positive definite. This assumption is equivalent to the "strengthened Legendre-Clebsche" necessary condition of the calculus of variations; therefore, the requirement that R be positive definite is sometimes referred to by that nomenclature.

In summary, a solution to the matrix Riccati equation (8) specifies, in the form of equations (9), a control which satisfies the necessary conditions (Euler-Lagrange equations) for optimality;

all that remains is to investigate sufficient conditions for the optimality of (9). The existence of a solution to the Riccati equation and an additional sufficient condition are related to the non-existence of conjugate points which are defined below.

The Conjugate Point Condition [G1]

Definition 2.1

A point t_1 is called a conjugate point to t_f if there exists a solution to the Euler-Lagrange equations (6) with the boundary conditions

$$x(t_1) = x(t_f) = 0, \text{ where } t_o \leq t_1 < t_f,$$

which is not zero everywhere on the interval.

Graphically, a conjugate point can be illustrated as in Figure 2.1 (cf. Figure 1 in Reference [B3]).

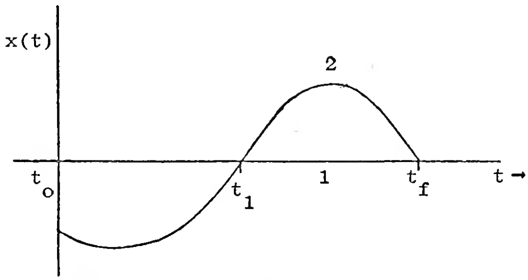


Figure 2.1 Conjugate Trajectory

Lemma 2.1 [B3]

When a conjugate point exists, both the trivial solution (Path 1 in Figure 2.1) and the conjugate trajectory (Path 2 in Figure 2.1) result in a zero value for the integral

$$\int_{t_1}^{t_f} (x^T Q x + u^T R u)\, dt.$$

Proof

Consider the identically zero integral on the conjugate path

$$\int_{t_1}^{t_f} \lambda^T(\dot{x} - Fx + GR^{-1}G^T\lambda)dt = 0$$

or
$$\int_{t_1}^{t_f} \lambda^T\dot{x}dt - \int_{t_1}^{t_f} \lambda^T(Fx - GR^{-1}G^T\lambda)dt = 0.$$

Applying integration by parts to the first term results in

$$\lambda^T x \Big|_{t_1}^{t_f} - \int_{t_1}^{t_f} \dot{\lambda}^T x + \lambda^T(Fx - GR^{-1}G^T\lambda)dt = 0.$$

The first term is clearly zero and with the substitution for $\dot{\lambda}$ from the Euler-Lagrange equations (6) and for u from (5)

$$\int_{t_1}^{t_f} (x^TQx + u^TRu)dt = 0,$$

which completes the proof.

When a conjugate point does exist at t_1, $\lambda(t_1)$ must be non-zero to be distinct from the identically zero solution to (6). If the system is controllable, it must respond to the input (5) resulting from $\lambda(t_1) \neq 0$; hence by continuity the conjugate path (λ and x) must be non-zero for some finite interval within (t_1,t_f). Then the conjugate path cannot be a duplicate of the identically zero path with the addition of a (non-zero) discontinuity; thus the presence of a conjugate point represents more than merely misbehavior at a single point. The occurrence of a conjugate point coincident with t_o is obviously disconcerting because it indicates that there are at least two candidates (that satisfy the necessary conditions) for an optimal control which lead

to entirely different trajectories at the same (zero) cost. When a conjugate point occurs at $t_1 \neq t_o$, the results are equally catastrophic but this case does not lend itself as well to heuristic interpretation.

Theorem 2.1

> For a linear, completely controllable system subject to a quadratic performance index (2) with R positive definite, a control u^* satisfying the Euler-Lagrange equations is globally optimal if and only if there exist no conjugate points on the interval (t_o, t_f).

A discussion of this theorem and its proof in a somewhat different context can be found in Breakwell and Ho [B3]. The requirement that conjugate points be non-existent is generally referred to as the "conjugate point condition" or occasionally in the classical calculus of variations as the Jacobi condition [G1].

Complete controllability requires that there exists an input, u(t), which will drive the system

$$\dot{x} = Fx + Gu(t) \ , \qquad x(t_o) = x_o$$

from any initial state, x_o, to the origin within an arbitrary time interval; this can be shown to be equivalent to requiring that the matrix $[G, FG, \ldots, F^{n-1}G]$ have full rank [K5]. Controllability is not actually a severe criterion in a practical sense. A plant which is not completely controllable can be transformed into two canonical subsystems, one containing the completely controllable part and the other subsystem containing the remainder of the plant dynamics [K5]. Hence, a given design problem can be thought of as two related designs on the canonical subsystems and it is only necessary to be certain that the non-controllable part is stable and/or not reflected in the performance index.

Now the solution to the linear, quadratic cost optimization problem is essentially complete; it is merely necessary to solve the Riccati equation for a candidate control law and to insure that the conjugate point condition is satisfied. Although there appears to be no convenient way to test for the presence of conjugate points, the following theorem demonstrates that a test for the conjugate point condition is actually implicit in the solution of the Riccati equation.

Theorem 2.2

> The solution to the Riccati equation (8) for the optimization problem specified by (1) (completely controllable) and (2) (with R positive definite) fails to exist (becomes unbounded) at t_1, $t_o \leq t_1 < t_f$, if and only if t_1 is a conjugate point to t_f.

A formal proof can be found in Lee [L2] which is very much in the spirit of the following heuristic justification.

From the definition of a conjugate point, $x(t_1) = 0$; however, $\lambda(t_1) \neq 0$, since this could only come about in the trivial solution. Recall that the solution to the Riccati equation was defined by (7) as

$$\lambda(t) = P(t,t_f)x(t).$$

Then as t approaches t_1, x approaches zero but λ remains non-zero; hence, $\|P(t)\|$ must correspondingly increase and finally become unbounded at $t = t_1$.

Thus far the conjugate point condition has been presented in a rather esoteric format with little physical interpretation. Again consideration of the general case is difficult but the occurrence of a conjugate point at the initial time (t_o) has an interesting interpretation. The absence of a conjugate point at t_o (in the linear

case) insures that no two system trajectories which satisfy the Euler-Lagrange equations (extremals) will ever intersect,[2] that is, there will never exist two distinct optimal trajectories emanating from a single system state [B4].

The solution to the Riccati equation (8) also has a physical interpretation which will be useful later.

Lemma 2.2

The value of the performance index for the optimal control law is

$$J^o(t_o, x_o, t_f) = \frac{1}{2}x_o^T P(t_o; t_f) x_o \tag{10}$$

where P is the solution to the matrix Riccati equation (8) which is bounded on (t_o, t_f), i.e., there exist no conjugate points.

Proof

$$J = \frac{1}{2}\int_{t_o}^{t_f} (x^T Q x + u^T R u)\,dt$$

Substitution of the optimal control law,

$$u = -R^{-1}G^T P x,$$

results in

$$J = \frac{1}{2}\int_{t_o}^{t_f} x^T(Q + PGR^{-1}G^T P)x\,dt. \tag{11}$$

---

[2] Intersection here is taken so as to exclude the case of tangential coincidence. In the general (non-linear) case extremals taken sufficiently close (i.e., neighboring trajectories) must not cross.

Define

$$F_k = F - GR^{-1}G^TP$$

the closed-loop state matrix and $\Phi_k(t_o,t)$ as its corresponding state transition matrix, i.e.,

$$\dot{\Phi}_k = F_k\Phi_k \quad \text{and} \quad x(t) = \Phi_k x_o.$$

Equation (11) can now be rewritten as

$$J = \frac{1}{2}\int_{t_o}^{t_f} x_o^T\Phi_k^T(Q + PGR^{-1}G^TP)\Phi_k x_o dt \ . \tag{12}$$

Substitution of the definition of F_k into the Riccati equation results in

$$\dot{P} = - F_kP - PF_k - PGR^{-1}G^TP - Q$$

and when employed in (12) to replace Q, (12) becomes

$$J = \frac{1}{2}\int_{t_o}^{t_f} x_o^T\Phi_k^T(-PF_k - F_k^TP - \dot{P})\Phi_k x_o dt$$

or

$$J = -\frac{1}{2}\int_{t_o}^{t_f} x_o^T\Phi_k^TP\dot{\Phi}_k x_o dt - \frac{1}{2}\int_{t_o}^{t_f} (x_o^T\dot{\Phi}_k^TP\Phi_k x_o + x_o^T\Phi_k^T\dot{P}\Phi_k x_o)dt.$$

Application of integration by parts to the first integral above results in terms which cancel the second integral, leaving

$$J = -\frac{1}{2} x_o^T\Phi_k^TP\Phi_k x_o \Big|_{t_o}^{t_f} = \frac{1}{2} x_o^TP(t_o,t_f)x_o - \frac{1}{2} x^T(t_f)P(t_f,t_f)x(t_f).$$

Since $P(t_f,t_f) = 0$, this is the desired result.

This proof is considerably different from the standard one (e.g., [A1, p. 25]) in that it does not require the use of

Hamilton-Jacobi theory. This lemma not only provides a necessary link in the solution of the problem considered in the next section but provides a tie with dynamic programming, in that formulation equation (10) is referred to as an "optimal return function" [B5].

Sufficient Conditions - Existence

The conjugate point condition is not entirely satisfactory from the standpoint of a working sufficiency condition. It is not known at the onset whether or not the optimization problem has a solution; only after the complete control law is computed is one assured that it was not all for naught. What is required then, it would seem, is a sufficient condition somewhat more restrictive than the conjugate point condition with the advantage that success of the optimization problem is predetermined.

Theorem 2.3

The optimization problem of minimizing a quadratic performance index

$$J = \int_{t_o}^{t_f} (x^T Qx + u^T Ru)dt, \quad R = R^T \text{ positive definite} \tag{13}$$

subject to a completely controllable linear plant

$$\frac{dx}{dt} = Fx + Gu \tag{14}$$

always has the global minimum

$$u(t) = - K^T(t)x(t) \quad \text{where} \quad K(t) = P(t,t_f)GR^{-1}$$

$$\text{and} \quad \frac{dP}{dt} = -F^T P - PF + PGR^{-1}G^T P - Q, \quad P(t_f,t_f) = 0 \tag{15}$$

for all finite (t_o, t_f) if Q is positive semidefinite and symmetric.

Proof

Suppose that Q is positive semidefinite and the Riccati equation (15) diverges at $t = t_1 < t_f$. A solution for (15) must exist in a sufficiently small neighborhood of t_f, further P(t) exists for all $t \in (t_1+\epsilon, t_f)$, $\epsilon > 0$; hence, by Lemma 2.2

$$J^o(x(t_1+\epsilon), t_1+\epsilon) = x^T(t_1+\epsilon)P(t_1+\epsilon, t_f)x(t_1+\epsilon) \geq 0, \qquad (16)$$
$$\text{for} \quad \epsilon < t_f - t_1$$

the inequality is due to the positive definiteness of the integrand of J. As ϵ approaches zero, some entry of P becomes unbounded. It can be assumed without loss of generality that at least one diagonal element of P becomes infinite; otherwise, some 2 x 2 principal minor of P would be negative, contradicting the positive semidefiniteness of P inferred in (16). Let e_i be a vector which is all zeros except for the ith element, a one, which corresponds to a diagonal term p_{ii} of P which becomes unbounded as ϵ approaches zero; then,

$$J^o(e_i, t_1+\epsilon) = p_{ii}(t_1+\epsilon, t_f).$$

Since J^o is the optimal performance index, a performance index resulting from any arbitrary control, say u = 0, must be greater. Then if $t_1 + \epsilon$ and e_i are chosen as the initial time and state of the optimization problem, i.e., $t_o = t_1 + \epsilon$,

$$J^o(e_i, t_1+\epsilon) = p_{ii}(t_1+\epsilon, t_f) \leq \int_{t_o}^{t_f} e_i^T \Phi^T(\tau-t_o) Q \Phi(\tau-t_o) e_i \, d\tau \qquad (\ 7)$$

where $\Phi(t-t_o)$ is the system state transition matrix and $\Phi(t-t_o)e_i$ is the resulting free trajectory. Clearly, the integral in (17)

remains bounded over the finite interval $(t_1+\epsilon, t_f)$, while the left-hand side becomes unbounded as $\epsilon \to 0$. This contradicts the original supposition of a conjugate point at t_1 and proves the theorem.

An interesting proof of this theorem using Lyapunov's second method to determine the stability of the Riccati equation is given in [K2].

Throughout the literature there appear sporadically references to theorems such as 2.3 as necessary and sufficient conditions for a solution of the optimization problem to exist (e.g., [D3,p. 557]); this is patently false and numerous counterexamples exist [G3]. No necessary and sufficient condition for the non-existence of conjugate points (other than integrating the Riccati equation) is presently known, although many researchers are actively pursuing these conditions and there is reason to believe they will be found in the near future [B6].

2.4 The Optimal Linear Regulator - Infinite Final Time

Existence and Stability

The Euler-Lagrange equations and the conjugate point condition apply as well to the case when the final time is no longer finite. However, as before, the conjugate point condition is not satisfactory as an existence test; the following theorem extends Theorem 2.3 to the infinite time case.

Theorem 2.4

If a linear plant (14) is completely controllable and a companion quadratic performance index,

$$J = \int_{t_o}^{\infty} (x^T Qx + u^T Ru)dt, \quad R = R^T \text{ positive definite} \tag{18}$$

has $Q = Q^T$ positive semidefinite and if $P(t,t_f)$ is a solution to the matrix Riccati equation (15) with $P(t_f,t_f) = 0$, then

$$\lim_{t_f \to \infty} P(t,t_f) = \bar{P}(t) \tag{19}$$

exists for all t and is a solution of (15).

Proof [K2]

First, it must be shown that the limit (19) exists for all t. Since the plant is completely controllable for every x_o, there exists a control $u'(t)$ which transfers x to 0 by some $t \leq t_1$. Set $u(t) = 0$ for $t > t_1$. Then

$$J^o(t_o,x_o,t_1) = x_o^T P(t_o,t_1)x_o \leq J(t_o,x_o,t_1)\Big|_{u'} = J(t_o,x_o,\infty)\Big|_{u'}$$

is bounded for all $t_1 > t_o$. The optimal costs are also non-decreasing as $t_1 \to \infty$. Suppose that this were not the case, that is, suppose

$$J_1^o(t_o,x_o,t_1) > J_2^o(t_o,x_o,t_2), \quad \text{for } t_2 > t_1,$$

where $u_1(t)$ is the optimal control corresponding to J_1^o and $u_2(t)$ is the optimal control resulting in J_2^o. Then by the positive definiteness of the integrand of (18), use of control u_2 will result in a lower cost at time t_1 than u_1, thus contradicting the optimality of u_1 and demonstrating the assertion that the optimal costs are monotonic. The limit therefore exists for all t by the well-known result that all bounded monotonic sequences possess a limit.

Now it is necessary to demonstrate that (19) is a solution to the Riccati equation. Define $P(t,t_f;A)$ as a solution of (15) with the boundary value $P(t_f,t_f) = A$. Then, using the continuity of solutions of (15) with respect to its boundary values,

$$\bar{P}(t) = \lim_{t_2 \to \infty} P(t,t_2;0) = \lim_{t_2 \to \infty} P(t,t_1;P(t_1,t_2;0))$$

$$= P(t,t_1;\lim_{t_2 \to \infty} P(t_1,t_2;0)) = P(t,t_1;\bar{P}(t_1))$$

and $\bar{P}(t)$ is a solution of (15) for all t.

The price that is paid for guaranteeing the existence of a solution over all time (in addition to requiring Q to be positive semidefinite) is the complete controllability of the plant.

<u>Corollary</u> <u>2.1</u>

The solution $\bar{P}(t)$ to the matrix Riccati equation (15) for the optimal linear regulator problem of Theorem 2.4 is constant and the unique positive definite solution of

$$\bar{P}F + F^T\bar{P} - \bar{P}GR^{-1}G^T\bar{P} + Q = 0, \tag{20}$$

the steady-state (algebraic) Riccati equation.

The stationarity of $\bar{P}$ follows from the arbitrariness of the choice of the initial time t_o, since all initial times must result in the same value of the optimal performance index. Equation (20) ensues when the effect on $\dot{\bar{P}}$ of the irrelevance of t_o is considered; the choice of the positive definite solution is dictated by the positive definiteness of (18) and uniqueness is guaranteed by the conjugate point condition.

It is now clear that the optimal linear regulator is the optimal control theory analog of the classical design using state feedback.

As pointed out by Kalman [K2], the literature contains many references where the stability of an optimal system is tacitly assumed. For example, Letov [Z1, p. 378] equates (without proof) optimal systems and those which are stable in the sense of Lyapunov; although probably true for the class of optimal systems which are also stable, it is not in general. This can be easily demonstrated with a simple example. Let a scalar input plant,

$$\dot{x} = Fx + gu,$$

have one or more eigenvalues with positive real parts and let the performance index to be minimized be,

$$J = \int_{t_o}^{t_f} u^2 dt,$$

that is, $Q = 0$ and $R = 1$. The problem so defined clearly has a solution (Theorem 2.3), which is

$$u(t) = 0.$$

Stability was not a result of optimization in this case primarily because the states are not reflected in the cost (performance index); had the plant been stable, however, the optimal system would have been also. The final theorem in this section formalizes this observation into a general result.

Theorem 2.5 [K1]

An optimal linear regulator problem satisfying the conditions of Theorem 2.3 results in an asymptotically stable control law if

i) the pair $[F,H]$, where $Q = HH^T$ is completely observable, i.e., $[H, F^T H, \dots (F^T)^{n-1} H]$ has full rank;

and only if

ii) the linear subspace

$$X_1 = \{x \neq 0 \mid \| He^{Ft} x \| = 0\}$$

of the state space is null (i) or $e^{Ft}x$, $x \in X_1$ is an asymptotically stable response in the sense of Lyapunov.

Proof

i) By the assumption of observability and the positive semi-definiteness of Q, the integrand of the performance index must be positive along any non-zero trajectory of the plant. Then

$$J^o(t,x(t)) = x^T(t)\bar{P}x(t) > 0, \quad x(t) \not\equiv 0 \tag{21}$$

along optimal trajectories. Differentiation of the performance index,

$$J(t,x(t)) = \int_t^\infty (x^T Qx + u^T Ru)\, d\tau,$$

results in $\dot{J}(t,x(t)) = -(x^T Qx + u^T Ru) < 0$, $x(t) \neq 0$, which is negative along all non-trivial trajectories. Then (21) is a Lyapunov function [L2] and the system is asymptotically stable by Lyapunov's Second Stability Theorem [L3, p. 37].

ii) When certain of the states are not observable in the cost, Lyapunov's method is effectively applied only to a subsystem (i.e., the

states observed in the cost) and the more general result follows when the stability of the remainder of the system is considered.

The requirements for stability placed on Q by Theorem 2.5 are not unduly restrictive in that they reflect good engineering judgment; that is, if a state has significant effect on system performance it should influence the design process. Some authors choose to include stability tacitly in the definition of the performance index by considering only the case where Q is positive definite (e.g., [S1]). Stability is assured since a positive definite Q must have a nonsingular factor and condition (i) of Theorem 2.5 is clearly satisfied.

2.5 Summary

Over the past ten years the optimal linear regulator problem has become one of the most widely studied optimal control problems. Perhaps the best way to summarize this unique problem is to review the qualities which set it apart.

1. The solution to the problem is always approached in the same fashion regardless of the specific system or weighting matrices employed and in contrast to the majority of optimal control problems, the solution is explicit.
2. In a restricted, but very large, set of weighting matrices (i.e., Q positive semidefinite) the solution is guaranteed to exist.
3. The solution is always in the form of a linear, constant feedback control law, as opposed to the general optimal control problem where a feedback formulation is not directly

obtainable. With another minor concession to generality (Theorem 2.5) asymptotic stability will be assured from the onset.

4. Numerical computation of the control laws is both straightforward and relatively inexpensive (in terms of computing time)[B7]; in addition, a multitude of numerical schemes are available [B1].

Under suitable assumptions of continuity and corresponding definitions of controllability and observability [D2] most of these remarks are also applicable to linear time-varying systems and non-constant weighting matrices [K2].

CHAPTER III

THE INVERSE OPTIMAL LINEAR REGULATOR PROBLEM

3.1 Introduction

Optimality in itself is not necessarily a desirable characteristic for a system to possess. For instance, suppose it is desired to minimize the sensitivity of an existing system's response to component variations by use of an appropriate controller. A system design which obviously has a minimal sensitivity is one that does nothing whatever; then one candidate for an optimal controller would be one that turns the system off. In this case, the optimal design would probably not be satisfactory because it was not implicit in the procedure that the resulting system should perform in some acceptable manner in addition to minimizing the performance index.

The lesson is clear, optimality may be a frivolous notion unless its ramifications are thoroughly understood in the context of total system behavior. In the present chapter the implications of optimality are considered for a single input linear plant subject to a quadratic performance index to set the stage for its use as a viable design tool.

The system to be considered in this chapter can be described by the state equation,

$$\frac{dx}{dt} = Fx + gu, \tag{1}$$

and occasionally it will require an associated feedback control law,

$$u = -k^T x, \tag{2}$$

where F is a constant n x n matrix and g,k are constant n-vectors. In conjunction with this plant, a quadratic performance index,

$$J = \int_0^\infty (x^T Q x + ru^2)\,dt, \tag{3}$$

will be studied. The control weighting factor, r, will be taken without loss of generality as unity throughout the remainder of this work.

3.2 When Is a Linear Control System Optimal?

The original impetus to investigate the "inverse optimal control problem" is generally considered to be a 1964 paper by R. E. Kalman, bearing the same title as this section [K1], although it has a much older history in the Calculus of Variations. In that paper, Kalman considers the theoretical criterion for a linear system to be optimal with respect to a restricted class of quadratic performance indices. His principal result is reviewed next.

Theorem 3.1 [K1]

Consider a completely controllable linear plant (1) with a stable completely observable control law k (2). Then k is an optimal control law with respect to a quadratic performance index (3), with Q positive semidefinite, if and only if,

$$\left|1 + k^T \Phi(j\omega) g\right|^2 \geq 1 \tag{4}$$

for all real ω, where $\Phi(s) = (sI - F)^{-1}$.

Proof

i) Necessity: If k is an optimal control law, then by Theorems 2.3 and 2.4

$$k = Pg \tag{5}$$

where P is a solution to the steady-state Riccati equation,

$$PF + F^T P - Pgg^T P + Q = 0. \tag{6}$$

Substituting (5) into (6) permits it to be rewritten as

$$PF + F^T P - kk^T + Q = 0. \tag{7}$$

Adding and subtracting sP to (7) results in

$$-P(sI - F) - (-sI - F^T)P - kk^T + Q = 0.$$

Premultiplying and postmultiplying the above by $g^T \Phi^T(-s)$ and $\Phi(s)g$, respectively (where $\Phi(s)$ is as defined in equation (4)) and substituting (5) where appropriate, leads to

$$g^T \Phi^T(-s)k + k^T \Phi(s)g + g^T \Phi^T(-s)kk^T \Phi(s)g = g^T \Phi^T(-s)Q\Phi(s)g$$

which may be factored as

$$[1 + g^T \Phi^T(-s)k][1 + k^T \Phi(s)g] = 1 + g^T \Phi^T(-s)Q\Phi(s)g. \tag{8}$$

Noting the semidefiniteness of Q proves the necessary condition.

ii) Sufficiency: Proof of the sufficient part of the theorem is not particularly enlightening and requires concepts that are yet to appear, hence it will be deferred to the appendix.

3.3 Implications of Optimality

It is somewhat surprising that the condition for optimality (4) should be a frequency domain criteria in that all of the previous

analysis of the problem (Chapter II) was carried out in the time domain. The study of optimal systems in the frequency domain may be responsible for the current interest shown by Kalman [K6] and others in realization theory from the point of view of invariants and thus promises to have far greater influence in years to come.

Examination of the optimality criterion (4) reveals that the quantity whose modulus is taken on the left-hand side of the equation, i.e.,

$$f_k(j\omega) = 1 + k^T \Phi(j\omega) g$$

is the "return difference" of classical feedback theory [B8]. It can be interpreted as the difference between a unity input (in the frequency domain) to the system and what would consequently be returned as feedback. Graphically, this is illustrated in Figure 3.1 as the difference between the input at node a and the feedback at node b.

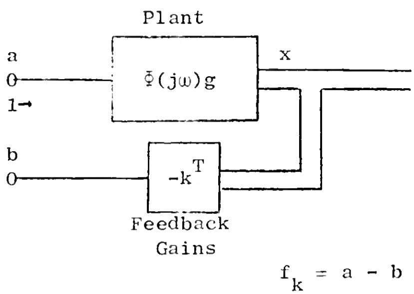


Figure 3.1 Return Difference

The requirement of (4) that the modulus of the return difference be greater than unity is the celebrated result of classical feedback theory that the sensitivity of system response to variations

in plant parameters is reduced by the addition of feedback [B8]. Further analysis of inequality (4) reveals that it requires that the Nyquist locus avoid a circle of unit radius about the point (-1,0). Figure 3.2 pictures a Nyquist plot of a hypothetical optimal system constructed from a plant with two eigenvalues with positive real parts.

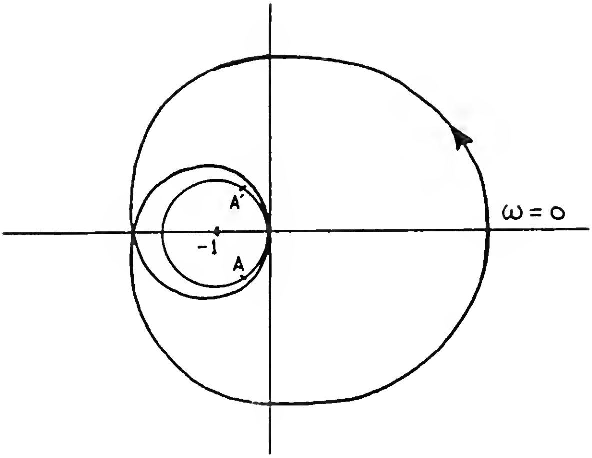


Figure 3.2 Nyquist Plot of Optimal System

Two additional observations may be immediately made from Figure 3.2. First, the phase margin of an optimal system must be at least 60^{0}, since the closest points (in phase) to the negative real axis that the locus can cross the unity magnitude contour (points A and A' on Figure 3.2) are displaced 60^{0} from the axis. Second, the system will clearly remain stable regardless of how the feedback gain is increased. Then if gain margin is defined as that factor by which feedback gain may be <u>increased</u> before system instability occurs [E2], the gain margin of an optimal system may be said to be infinite. However, if gain margin is defined as the reciprocal of the magnitude of the loop gain at phase crossover, i.e.,

$$GM = \frac{1}{|k^T\Phi(j\omega_c)g|} \quad \text{where} \quad \text{Arg}\,[k^T\Phi(j\omega_c)g] = 180°,$$

then an optimal system has a gain margin of at least $\frac{1}{2}$ or approximately -6 db. Kuo indicates that the two definitions given are equivalent [K7, p. 398], but consideration of the example of Figure 3.2 reveals that this is not the case when the plant is non-minimum phase.

A third definition, and the one that should be understood in any references to gain margin in the sequel, is the following [T2]:

> Let g_i be the factor by which the feedback gain may be increased until instability occurs and g_d be the factor by which the gain may be decreased and stability maintained, then
>
> $$\overline{GM} = \text{smaller of } \{g_i, 1/g_d\}.$$

This definition is more reasonable in that it reflects the actual gain disturbance required for instability and is easily determined from the Nyquist diagram. The gain margin for an optimal regulator, in the sense of Theorem 3.1, then is at least 6 db.

It is instructive at this point to compare these figures of gain and phase margin with those required in practice. Jones, Moore and Tecosky in Truxal's classic handbook [T2, pp. 19-14] recommend phase margins of 40° to 60° and gain margins of at least 10 db for applications typical of chemical process control. Generalities of this sort are of course subject to severe criticism but nonetheless it can be concluded that minimum optimal regulator specifications compare favorably with those desired of classical designs.

3.4 Companion Matrix Canonical Form

It is well known that for any non-derogatory matrix, F, there exists a similarity transformation, T, such that

$$\hat{F} = TFT^{-1}$$

is the companion matrix [M1] of F. A companion matrix being a matrix of the form:

$$\hat{F} = \begin{bmatrix} 0 & 1 & 0 \ldots 0 \\ 0 & 0 & 1 \ldots 0 \\ & & \ldots & \\ -a_1 & -a_2 & -a_3 \ldots -a_n \end{bmatrix} \tag{9}$$

where

$$\varphi(s) = s^n + \sum_{k=1}^{n} a_k s^{k-1} = \det\,(sI - F).$$

That is, the last row of the companion matrix of F contains the negative of the normalized coefficients of the characteristic polynomial of F (and $\hat{F}$), with the remainder of the matrix null save a superdiagonal of ones. Some authors choose to refer to the transpose of the matrix defined above as the companion matrix [G4] In the following either definition will suffice; however, (9) will be assumed for consistency.

When F assumes the role of the state matrix in a completely controllable linear system (1) the similarity transformation, T, which transforms F to its companion matrix, places the system as a whole in a canonical form. Specifically, if z = Tx represents the change of basis described above and the linear completely controllable system is

$$\dot{x} = Fx + gu$$

$$y = H^T x,$$

then [K1]

$$\dot{z} = \hat{F}z + \hat{g}u$$

$$y = \hat{H}^T z$$

$$\hat{F} = TFT^{-1}, \text{ the companion matrix of } F$$

$$\hat{g} = Tg = \begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{matrix}, \text{ and } \hat{H} = (T^{-1})^T H.$$

Complete controllability allows the specified form of $\hat{g}$ to be chosen. Since there is no loss of generality, controllable systems will be assumed in this coordinate system and the "hat" notation will be omitted.

The problem of actually computing the required transformation has lately occupied a good deal of the literature, e.g., [J1,R2,R3]. Computation of the transformation is conceptually not very difficult and some straightforward numerical solutions have been formulated [esp. R3].

The companion matrix canonical form is often referred to as the phase-variable canonical form [S2]; this cognomen alludes to the property that each state is the derivative of the preceding state. A second property that will find application is the fact that g has only one non-zero element, hence, the control directly affects only the last state (highest derivative).

3.5 Characterization of the Equivalence Class of Q's

Employment of the coordinate system reviewed in the last section permits additional insight to be gleaned from Theorem 3.1. First it will be necessary to exploit a characteristic of the companion matrix canonical form.

Lemma 3.1

For a completely controllable linear system in the companion matrix canonical form (1)

$$S(s) = \varphi(s)\Phi(s)g = \begin{bmatrix} 1 \\ s \\ s^2 \\ \vdots \\ s^{n-1} \end{bmatrix},$$

where $\Phi(s) = (sI - F)^{-1}$
and $\varphi(s) = \det\,(sI - F)$.

Proof

To prove the theorem it is sufficient to show that

$$(sI-F)\varphi(s)\Phi(s)g = (sI-F)S(s) = \varphi(s)g.$$

With F in companion matrix form

$$(sI-F)S(s) = \begin{bmatrix} s & -1 & 0 & \dots & 0 \\ 0 & s & -1 & \dots & 0 \\ & & \cdot & \dots & \\ a_1 & a_2 & a_3 & \dots & a_n+s \end{bmatrix} \begin{bmatrix} 1 \\ s \\ \vdots \\ s^{n-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ s^n + \sum_{i=1}^{n} a_i s^{i-1} \end{bmatrix}.$$

which is clearly $\varphi(s)g$ and the theorem is proven.

This lemma allows the simplification of equation (8) used in the proof of Theorem 3.1 to

$$\left| 1 + \frac{k^T S(j\omega)}{\varphi(j\omega)} \right|^2 = 1 + \frac{S^T(-j\omega) Q S(j\omega)}{|\varphi(j\omega)|^2} . \qquad (10)$$

Multiplying by $|\varphi(j\omega)|^2$ results in

$$|\varphi(j\omega) + k^T S(j\omega)|^2 = |\varphi(j\omega)|^2 + S^T(-j\omega)QS(j\omega).$$

With the observation that the expression inside the absolute value marks is the closed-loop characteristic polynomial (which will be denoted $\varphi_k(s)$) permits further simplification to

$$|\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2 = S^T(-j\omega)QS(j\omega). \tag{11}$$

This relation (11) defines a polynomial by which the open- and closed-loop characteristic equations of an optimal system must be related to the state weighting matrix of a corresponding performance index. This polynomial will appear frequently; consequently, it will be convenient to refer to it with the following notation:

$$\Psi(\omega) = |\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2 = S^T(-j\omega)QS(j\omega), \tag{12}$$

or occasionally as

$$\Psi(Q;\omega) = S^T(-j\omega)QS(j\omega)$$

to emphasize the functional relationship of Q.

Recalling that characteristic polynomials are invariant under similarity transformation and the assumption of positive semidefiniteness for Q leads to a useful corollary of Theorem 3.1.

Corollary 3.1

A completely controllable scalar linear system (1) with a completely observable stable control law k is optimal with respect to a quadratic performance index (2) with Q positive semidefinite and completely observable if and only if

$$|\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2 \geq 0 \quad \text{for all real } \omega,$$

where $\varphi(s) = \det(sI - F)$ and $\varphi_k(s) = \det(sI - F + gk^T)$, the open- and closed-loop characteristic polynomials, respectively.[1]

The requirement that the feedback control be completely observable again appears without apparent justification. Suppose that k were not completely observable, then the pair of polynomials which comprise the expression

$$k^T \Phi(j\omega) g = \frac{K(j\omega)}{\varphi(j\omega)}$$

will have a common factor [K5]. This can be thought of, in the classical control sense, as having a zero on top of a pole which prevents the response of the pole from being observed at the output. Let the common factor be $\gamma(j\omega)$ and denote the polynomials with γ factored with a prime, then equation (10) becomes

$$\left| 1 + \frac{K'(j\omega)\gamma(j\omega)}{\varphi'(j\omega)\gamma(j\omega)} \right|^2 = 1 + \frac{S^T(-j\omega)QS(j\omega)}{|\varphi'(j\omega)|^2 |\gamma(j\omega)|^2}$$

or

$$|\varphi'(j\omega) + K'(j\omega)|^2 |\gamma(j\omega)|^2 = |\varphi'(j\omega)|^2 |\gamma(j\omega)|^2 + S^T(-j\omega)QS(j\omega).$$

The implication is that $S^T(-j\omega)QS(j\omega)$ must contain the common factor $|\gamma(j\omega)|^2$ as well. Since Q is positive semidefinite, it can be factored as $Q = HH^T$; then by the preceding statement the vector of polynomials formed from

---

[1] Through a minor abuse of the language, the terms "completely observable Q" and "observability from the cost" should be taken to mean that the pair [F,H] are completely observable when H is any factor of Q such that $Q = HH^T$. Similarly, "completely observable k" means that the pair [F,k] are completely observable.

$$H^T\Phi(j\omega)g \tag{13}$$

may have the common factor $\gamma(j\omega)$ and the Q may not be completely observable. The reason for the uncertainty in the observability of Q is the ambiguity in the location of the zeros of the factor of (13) corresponding to $|\gamma(j\omega)|^2$. Depending on the choice of H, (13) can have zeros in the right or left half-plane which may or may not eclipse a pole of the system. If k is completely observable, however, there will be no common factor regardless of how H is chosen and any Q which satisfies (13) will be completely observable. This is best illustrated with an example.

Example 3.1

Consider the second order plant,

$$F = \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix} \quad g = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

with the feedback law,

$$k = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The feedback is clearly not observable and in fact cancels one of the plant poles at -1 as shown in Figure 3.3.

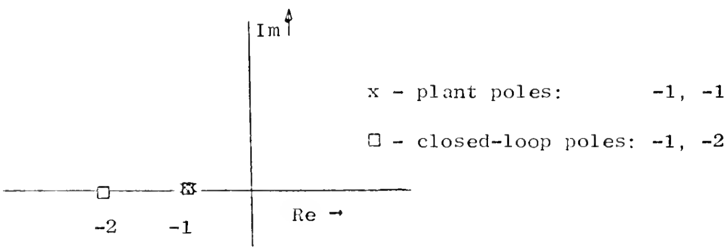


Figure 3.3 System Pole Locations

The control law is nonetheless optimal and three performance indices which are minimized by it have Q's,[2]

$$Q_1 = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \qquad h_1 = \sqrt{3}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} 3 & -3 \\ -3 & 3 \end{bmatrix} \qquad h_2 = \sqrt{3}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

and

$$Q_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}.$$

The first is not observable; (13) has a zero at -1 which coincides with a plant pole. The second is identical to Q_1 except that the zero of (13) is reflected about the axis (at + 1) and is consequently observable. The last is non-singular and avoids the difficulty entirely. The important point to note is that the control is optimal for all of these Q's. Complete observability from the cost is not required for optimality; it is only necessary that any plant poles which are unobservable appear in the closed-loop system as well. That is, optimal feedback cannot move system poles which are unobservable in the cost.

The requirement that k be completely observable is, in a sense, an "inverse" sufficient condition to the observability of Q; if the condition is met, then any Q for which the system is optimal must be completely observable. If, however, k is not completely observable but Ψ is non-negative, there may still exist one or more positive semidefinite Q's which are completely observable for which the system is optimal.

---

[2] These statements will be justified later.

Another corollary to Theorem 3.1 results from consideration of the preceding remarks.

Corollary 3.2

A completely controllable scalar linear system with a stable feedback control law is optimal with respect to a quadratic performance index with Q positive semidefinite if

$$|\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2 \geq 0 \qquad \text{for all real } \omega,$$

and

a) $\varphi(s)$ and $\varphi_k(s)$ are relatively prime which further insures that any Q for which the control law is optimal is completely observable,

or

b) the plant is stable.

Physically, the (a) condition can be interpreted as meaning that the aforementioned observability difficulties do not arise if the feedback moves all the plant poles. This is the case because the common factor in the discussion preceding the corollary does not exist. The second (b) condition removes the ambiguity by restricting the plant poles to the left-half complex plane (including the imaginary axis) and no difficulties of the sort discussed will occur. Some necessary and sufficient conditions similar to those of Corollary 3.2 will be considered coincidentally with later results.

In order to provide greater insight into the optimality condition in a strictly mathematical sense, it will be convenient to consider the requirements for Q implied by equation (11). In the same fashion that the two preceding corollaries dealt with the left-hand side of the equation and its relation to the system, it is desirable to consider what the right-hand side, i.e.,

$$\Psi(Q;\omega) = S^T(-j\omega)QS(j\omega),$$

portends for the corresponding performance index.

Before proceeding, two well-known lemmas on factorization will be stated; the first concerns a factorization of real, even polynomials and the second a factorization of positive semidefinite matrices.

Lemma 3.2

If and only if $\Gamma(\omega)$ is a real, even polynomial and

$$\Gamma(\omega) \geq 0 \qquad \text{for all real } \omega,$$

then there exists a unique "spectral" factor $\gamma(s)$ such that

$$\Gamma(\omega) = \gamma(j\omega)\gamma(-j\omega)$$

and $\gamma(s)$ has only zeros with non-positive real parts.

Many proofs of this lemma are recorded in the literature; for example, see Brockett [B2, p. 173 ff].

Lemma 3.3

A real, symmetric matrix Q may be factored as

$$Q = HH^T,$$

where H is a real matrix of rank (H) = rank (Q) if and only if Q is positive semidefinite.

This is a fundamental theorem which is encountered in the study of quadratic forms; see [A2, p. 139].

The groundwork has now been laid to present a theorem which concisely places the system-theoretic results of Theorem 3.1 into a mathematical framework.

Theorem 3.2

For every real, even polynomial $\Gamma(\omega)$ of order $2(n-1)$, there exists a real, symmetric, positive semidefinite, nth order matrix Q such that

$$\Psi(Q;\omega) = \Gamma(\omega),$$

(Ψ as defined in (12)) if and only if,

$$\Gamma(\omega) \geq 0 \qquad \text{for all real } \omega.$$

Proof

i) Sufficiency: If $\Gamma(\omega)$ is real, even and non-negative, Lemma 3.2 assures that a real factor $\gamma(s)$ exists such that

$$\gamma(j\omega)\gamma(-j\omega) = \Gamma(\omega). \tag{14}$$

Let h be a real vector composed of the coefficients of $\gamma(s)$ ordered with the constant term first and the coefficient of the (n-1)st term last. Then, clearly,

$$\gamma(j\omega) = h^T S(j\omega) \quad \text{and} \quad \gamma(-j\omega) = S^T(-j\omega)h,$$

where S is as defined in equation (12) of Lemma 3.2. The product (14) can be identified as

$$\Psi(hh^T;\omega) = S^T(-j\omega)hh^T S(j\omega) = \gamma(-j\omega)\gamma(j\omega) = \Gamma(\omega)$$

and hh^T provides a positive semidefinite Q which satisfies the sufficiency part of the theorem.

ii) Necessity: It must be shown that any positive semidefinite Q results in a $\Psi(Q;\omega)$ that, is a real, even, non-negative polynomial. Since Q is positive semidefinite, the Hermitian form,

$$\Gamma(\omega) = \Psi(Q;\omega) = S^T(-j\omega)QS(j\omega),$$

is non-negative and is clearly real and even. The theorem is proved.

Theorem 3.2 relates the optimality condition of Corollary 3.1 to a realizability condition for positive semidefinite Q's. Corollary 3.1 states that if the polynomial

$$\Psi(\omega) = |\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2$$

is non-negative and the feedback control law is stable, the closed-loop system is optimal. Theorem 3.2 states that if an arbitrary real even polynomial is non-negative, then there exists a positive semidefinite matrix Q that generates the polynomial by equation (12). If the system

is in companion matrix canonical form, the polynomial studied for optimality and the polynomial of Theorem 3.2 are one and the same as revealed by equation (11). Together they constitute all that is required to determine the optimality of a given closed-loop system configuration with respect to a positive semidefinite Q and to construct a corresponding performance index.

The proof of Theorem 3.2 provides a "recipe" for computing at least one performance index which is minimized by a given canonical optimal system; that is, by spectral factorization. This realization also prevents the performance index from obscuring unstable system poles and thus avoids the observability problem. At this point, it would be instructive to return to Example 3.1 and verify system optimality and the choices of Q.

Example 3.2

Consider the second order plant and feedback law of Example 3.1,

$$F = \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}, \qquad k = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The open- and closed-loop characteristic polynomials are

$$\varphi(s) = s^2 + 2s + 1,$$

$$\varphi_k(s) = s^2 + 3s + 2,$$

and

$$\Psi(\omega) = |\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2 = 3\omega^2 + 3,$$

is clearly non-negative. Then since the plant is stable, the optimality of the feedback law follows directly from Corollary 3.2b. Theorem 3.2 insures that a positive semidefinite Q can be found which

will form a performance index minimized by the control law, and the proof hints at how one such Q can be constructed. Following the proof the spectral factor of Ψ is computed, i.e.,

$$\psi(s) = \sqrt{3}(s+1) \quad \text{and}$$

$$\Psi = \psi(-j\omega)\psi(j\omega).$$

Now a vector h is constructed composed of the coefficients of $\psi(s)$ ordered with the constant term first,

$$h = \sqrt{3}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad Q = hh^T = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix},$$

which is the choice for Q_1 in Example 3.1.

In order to avoid an increasing awkwardness in notation, the following definition is required.

<u>Definition 3.1</u>

A state weighting matrix Q_2 is said to be <u>equivalent</u> to another weighting matrix Q_1 if for a given linear system (1) the quadratic performance indices (3) formed from Q_1 and Q_2 are minimized by the same control law. This equivalence will be denoted by a tilde, i.e.,

$$Q_1 \sim Q_2,$$

and the set of all matrices which are equivalent for a given system and optimal control law will be referred to as the "equivalence class of Q's" for that optimal system.

Note that this equivalence relation is correctly defined in the strictest sense [A3], that is, it is:

i) reflective - $Q_1 \sim Q_1$

ii) symmetric - if $Q_1 \sim Q_2$, then $Q_2 \sim Q_1$

iii) transitive - if $Q_1 \sim Q_2$ and $Q_1 \sim Q_3$, then $Q_1 \sim Q_3$.

The definition fails to draw a distinction between symmetric and non-symmetric matrices. Consistent with the remainder of this work, symmetric weighting matrices will be tacitly assumed. There is no loss of generality in the assumption of Q symmetric; if A is a non-symmetric matrix, then it is easy to see that it can be replaced in a quadratic form by the symmetric matrix $\frac{1}{2}(A + A^T)$ without altering the value of the form [H1].

The next theorem, which may be considered a central result of this section, prescribes how <u>all</u> quadratic performance indices minimized by a given system are related.

<u>Theorem 3.3</u>

For a scalar system, $\dot{x} = Fx + gu$, in companion matrix canonical form with a stable optimal control law k, a state weighting matrix Q_2 is equivalent to a weighting matrix Q_1 which forms a quadratic performance index minimized by k_1 if and only if

a) $$\Psi(Q_2;\omega) = \Psi(Q_1;\omega),$$

b) the linear subspace of the state space,

$$X_1 = \{x \neq 0 \mid x^T e^{F^T t} Q_2 e^{Ft} x = 0\}$$

is null or $e^{Ft}x$ for $x \in X_1$ is an asymptotically stable response in the sense of Lyapunov, and

c) the optimization problem for Q_2 and the given system has no conjugate points on $t \in [0,\infty)$.

<u>Proof</u>

i) Sufficiency: By Theorem 2.1 and (c) above, the minimization of the performance index with Q_2 results in a unique optimal control law which by Theorem 2.5 and (b) is also stable; denote this control law as k_2. Suppose that $k_2 \neq k_1$, where k_1 is the optimal control law

corresponding to Q_1. Then, since the coefficients of the closed-loop characteristic polynomials are the sum of the plant characteristic polynomial coefficients and the entries of k_1 (or k_2),

$$\varphi_{k_2}(s) \neq \varphi_{k_1}(s),$$

and by the stability of k_1 and k_2

$$|\varphi_{k_2}(j\omega)|^2 \neq |\varphi_{k_1}(j\omega)|^2$$

(i.e., both characteristic equations have zeros in the left half-plane). This last inequality with the definition of Ψ (12) contradicts part (a) of the hypothesis and sufficiency is demonstrated.

ii) Necessity: It must now be shown that if $Q_2 \sim Q_1$, then (a), (b), and (c) follow. Parts (b) and (c) are implicit in that they are necessary and sufficient conditions for stability and existence, respectively, of the optimization problem with Q_2. The first part results from the equality of k_1 and k_2 and the definition of Ψ (12).

This theorem could easily have been rewritten to include the case where the resulting control law was not stable; however, this complication would have no usefulness and would obscure insight into the mechanism of equivalent Q's. Many of the results to follow can be extended to include unstable optimal control laws but any apparent increase in relevance is purely artificial. Hence the praxis of considering control laws to be only stable will be continued throughout the sequel.

As is often the case, in this theorem practicality is the price of generality. In the discussion of Theorem 2.3 it was observed that the conjugate point condition was not very satisfactory as a sufficiency

condition for the optimization. The same remarks apply here: no conjugate points will exist for Q_2 positive semidefinite; then the conjugate point condition need only be examined where Q_2 is not positive semidefinite. The condition that Ψ be identical for equivalent Q's, part (a), can also be simplified.

The next lemma provides the last link in reformulating the results of Theorem 3.3 into a workable equivalence relation for performance indices.

Lemma 3.4

For an arbitrary real, symmetric, nth order matrix Q, the real even polynomial resulting from the quadratic form

$$\Psi(Q;\omega) = S^T(-j\omega)QS(j\omega) \tag{15}$$

where

$$S(s) = \begin{bmatrix} 1 \\ s \\ \vdots \\ s^{n-1} \end{bmatrix}$$

can be determined from the relation

$$\Psi(Q;\omega) = \sum_{i=1}^{n} [q_{ii} + 2\sum_{j=1}^{i-1} (-1)^j q_{i-j,i+j}]\omega^{2(i-1)}. \tag{16}$$

($q_{i,j} = 0$ for $i < 1$ or $j > n$).

Proof

Rewrite the vector S as the sum of two orthogonal vectors,

$$S(s) = a(s) + b(s) \quad a(s) = \begin{bmatrix} 1 \\ 0 \\ s^2 \\ 0 \\ s^4 \\ \vdots \end{bmatrix} \quad b(s) = \begin{bmatrix} 0 \\ s \\ 0 \\ s^3 \\ 0 \\ \vdots \end{bmatrix}$$

and clearly

$$S(-s) = a(s) - b(s)$$

since a is an even function of s and b is odd. Substitution of this relation for S into (15) leads to

$$\begin{aligned}\Psi(Q;s) &= (a(s) - b(s))^T Q(a(s) + b(s)) \\ &= a^T Qa - b^T Qb + (a^T Qb - b^T Qa).\end{aligned}$$

By the symmetry of Q the bilinear terms in parentheses total zero.

Since a(s) and b(s) have odd or even elements respectively zero, $\Psi(Q;s)$ contains no terms with $q_{i,j}$ with i + j odd. The non-zero off-diagonal terms resulting from $a^T Qa$ have i and j both even, while the non-zero off-diagonal terms from $b^T Qb$ have i and j both odd. Hence, Ψ may be rewritten as

$$\Psi(Q;s) = \sum_{i=1}^{n} (-1)^{i+1} q_{ii} s^{2(i-1)} + \sum_{i,j \text{ odd}} q_{ij} s^{i+j-2} + \sum_{i,j \text{ even}} q_{ij} s^{i+j-2}$$

which reduces easily to (16).

This lemma is important in its own right in that it provides an algorithm for computation of Ψ for a given Q without the necessity of evaluating the quadratic form. Its primary value however is that it allows the equivalence condition of Theorem 3.3 to be redefined from the invariance of a polynomial (Ψ) to specific algebraic constraints on the entries of the matrices in question.

Corollary 3.3

A weighting matrix Q_2 is equivalent to a matrix Q_1 for the system of Theorem 3.3 if and only if the entries of the matrices are related so that the quantities

$$p_i = q_{ii} - 2q_{i-1,i+1} + 2q_{i-2,i+2} - \cdots \qquad (17)$$

$$i = 1,2,3\ldots,n$$

$$(q_{ij} = 0 \text{ for } i < 1 \text{ or } j > n)$$

are equal for q_{ij} taken to be elements Q_1 or Q_2 and parts (b) and (c) of Theorem 3.3 are satisfied.

Proof

It must be shown that the relation (17) given in the corollary is equivalent to part (a) of Theorem 3.3, that is, if it is satisfied, the Ψ polynomials for Q_1 and Q_2 are identical. This is accomplished by demonstrating that the coefficients of the respective Ψ polynomials are coincident. The proof follows directly from determination of the coefficients of Ψ from (16) which are then related to the p_i of (17).

With additional study of equations (16) and (17) a somewhat startling phenomenon comes to light. The only entries of a given weighting matrix Q which influence the polynomial $\Psi(Q;\omega)$ are of the form

$$q_{kk} \text{ and } q_{k-\ell,k+\ell},$$

which excludes any element q_{ij} where $i+j$ is odd. This indicates that approximately half of the elements of an arbitrary weighting matrix are irrelevant (assuming that parts b and c of Theorem 3.3 are still satisfied) with respect to the optimal control law. Kalman and Englar [K3] noted from the structure of the companion matrix canonical form that the q_{ij} elements where $i+j$ is odd are "irrelevant" without identifying the underlying structural relation (17) for equivalent Q's.

Their suggestion that the "irrelevant" terms in a given weighting matrix be nulled at the onset of the optimization procedure, in order to simplify computation, is potentially hazardous. There is the possibility that nulling these elements will alter the observability qualities of the original matrix to the extent that an unstable plant pole is obscured. Fortunately, this occurrence appears to be extremely unlikely; the author was unable to construct a matrix which behaved in this fashion after a very exhaustive search. It seems that an unobservable matrix will remain so after the $i + j$-odd terms have been struck and conversely an observable matrix will still be observable after these terms have been removed. A very convincing heuristic argument can be made in support of this observation; first, however, it will be convenient to state two lemmas.

Lemma 3.5 (Gerschgorin, 1931 [G5])

All the eigenvalues of a square matrix $A = (a_{ij})$ lie in the union of circular regions,

$$|a_{ii} - z| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \quad i = 1,2,3,\ldots,n,$$

of the complex plane.

This is, of course, the touted Gerschgorin Circle Theorem which has found wide application in the numerical eigenvalue problem. A laconic proof of this famous theorem can be found in Cullen [C2, p. 197]. Of principal immediate interest, however, will be a second lemma which, although a corollary of Gerschgorin's Theorem, has found prominence as a separate result.

Lemma 3.6

Diagonally dominant matrices are non-singular. A square matrix A is said to be diagonally dominant if

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \quad \text{for} \quad i = 1,2,3,\ldots,n \tag{18}$$

where a_{ij} is the (i,j)th element of A; that is, if the diagonal entries are larger in modulus than the sum of the magnitudes of the remaining constituents of their respective rows.

The justification for this lemma follows easily from Lemma 3.5. If a matrix is diagonally dominant, the region of permitted eigenvalue locations excludes the origin and the matrix is consequently non-singular.

Consider Figure 3.4 which is a schematic drawing of a symmetric matrix with the $i + j$ odd terms removed and the X's representing the remaining entries.

$$Q = \begin{bmatrix} X & 0 & X & 0 & X \ldots \\ 0 & X & 0 & X & 0 \ldots \\ X & 0 & X & 0 & X \ldots \\ & & \ldots & & \end{bmatrix}$$

Figure 3.4 Sparse Equivalent Matrix

It is clear from observation of the figure that removing the indicated entries tends to increase the dominance of the diagonal and in a sense makes it "more non-singular"; thus it would be reasonable to expect an aggrandizement of observability rather than a deterioration.

A possible second concern is that discarding elements in the manner described may destroy the positive semidefiniteness of the

weighting matrix. A second application of Gerschgorin's Theorem, this time in the form of Lemma 3.5, reveals that the likelihood of degrading the positive semidefiniteness of a Q matrix by removing the irrelevant terms is extremely small.

If the matrix is positive semidefinite, the union of circles which form the permissible regions for eigenvalues must include parts of the right-half complex plane (including, perhaps, the origin); and by a well-known result of matrix theory [H1], the diagonal elements of a positive semidefinite matrix, hence the centers of these circles, must be non-negative. Then as the off-diagonal terms are removed, as in Figure 3.4, it is apparent from relation (18) that the radii of the circles which may contain eigenvalues are reduced and the eigenvalues will consequently be restricted to fall in a region that is, if anything, more positive. The only case which is not resolved by this argument is the one where the permissible region of the original matrix includes part of the left half-plane and reduction of the radii does not retrieve the region wholly into the right half-plane; thus admitting the possibility of a negative eigenvalue in the reduced matrix. In any case, the migration of any of the eigenvalues of a Q matrix to the left, as a result of the simplifying operation, appears extremely unlikely.

In general the observability and absence of conjugate points for the sparser matrix must be tested. However, for a specific but very important case, both conditions (and hence equivalence) can be guaranteed a priori.

Theorem 3.4

If a weighting matrix Q_1 for a quadratic performance index (3) operating on a scalar system in companion matrix canonical form is unity rank, i.e.,

$$hh^T = Q_1,$$

where h is a vector, and the resulting optimal control law is stable, then the matrix Q_2 resulting from nulling the entries q_{ij} of Q_1 for $i+j$ odd is equivalent to Q_1.

Proof

Since Q_1 is unity rank, its elements can easily be written as functions of the elements of the factor vector; that is,

$$Q_1 = hh^T = \begin{bmatrix} h_1^2 & h_1h_2 & h_1h_3 & h_1h_4 & \cdots \\ h_1h_2 & h_2^2 & h_2h_3 & h_2h_4 & \cdots \\ h_1h_3 & h_2h_3 & h_3^2 & h_3h_4 & \cdots \\ & & \cdots & & \end{bmatrix},$$

where h_i is the ith entry of h. The matrix Q_2 constructed from Q_1 by discarding the elements q_{ij} where $i+j$ is odd is

$$Q_2 = \begin{bmatrix} h_1^2 & 0 & h_1h_3 & 0 & \cdots \\ 0 & h_2^2 & 0 & h_2h_4 & \cdots \\ h_1h_3 & 0 & h_3^2 & 0 & \cdots \\ & & \cdots & & \end{bmatrix}.$$

It is easily verified that a rank two factor of Q_2 is

$$H_2 = \begin{bmatrix} h_1 & 0 \\ 0 & h_2 \\ h_3 & 0 \\ 0 & h_4 \\ \vdots & \vdots \end{bmatrix}, \quad Q_2 = H_2H_2^T.$$

Part (a) of Theorem 3.3, i.e., $\Psi(Q_1;\omega) = \Psi(Q_2;\omega)$, is satisfied by virtue of Corollary 3.3 and part (c) is fulfilled because Q_2 is shown to be positive semidefinite as a result of the existence of the factor H_2. Then all that is required is to show that no unstable poles will be unobservable by H_2 (part b) to prove $Q_2 \sim Q_1$. In fact, an even stronger condition will be shown; that is, the only plant poles which are unobservable through Q_2 are precisely those which are unobservable through Q_1. This will effectively satisfy part (b) since Q_1 is required by the hypothesis to result in a stable control law and hence cannot fail to observe any unstable plant poles.

It must now be demonstrated that the only common factors of the polynomial entries of the vector,

$$H_2^T S(s), \tag{19}$$

are also factors of the corresponding polynomial for Q_1,

$$h^T S(s) \ .$$

If H_2 is thought of as defining two outputs to the system, the above requires that the outputs not simultaneously obscure any system pole which is observable through Q_1. The polynomials formed by (19) are

$$h_1 + h_3 s^2 + h_5 s^4 + \ldots$$

and

$$h_2 s + h_4 s^3 + h_6 s^5 + \ldots \ .$$

These polynomials can be recognized as the result of "separating" the polynomial $h^T S(s)$ into even and odd functions of s. This is a common procedure in Network Theory in which the polynomial operators Ev (·) and Od (·) are often defined [T3] to simplify notation.

In this notation

$$F(s) = h^T S(s) = Ev(F) + Od(F) = Ev(F) + sEv_o(F),$$

where $Ev_o(F)$ is $Od(F)$ with an s factored out. Making use of the property of even polynomials that their zeros lie symmetrically about the imaginary axis and denoting the highest even power coefficient of F as c_e and the highest odd power coefficient as c_o, F(s) can be rewritten as

$$F(s) = c_e \prod_{i=1} (s^2 + a_i) + s\, c_o \prod_{j=1} (s^2 + b_j).$$

In the above $\pm\sqrt{a_i}$ and $\pm\sqrt{b_j}$ are zeros of Ev(F) and $Ev_o(F)$, and the upper limits on the products must be chosen with regard to whether the dimension of the system is even or odd. Now if Ev(F) and $Ev_o(F)$ possess a common factor, then $a_k = b_m$ for some k and m and

$$F(s) = (s^2 + a_k)\left[c_e \prod_{\substack{i=1\\ i\neq k}} (s^2+a_i) + c_o \prod_{\substack{j=1\\ j\neq m}} (s^2+b_j)\right]$$

which is clearly a factor of F(s) as well, and the theorem is proven.

This proof of the observability of the rank two Q will be useful in the next theorem. A simpler analysis follows easily from consideration of the composition of the two outputs defined by H_2. Summing these outputs results in a single output identical to the one defined by h, thus any response which is observable from h is also observable from H_2.

This theorem guarantees that if a system and an optimal control minimize a single output in the mean-square sense (h is unity rank), then with little effort a pair of outputs can be defined which are also minimized. The real value of this result is, however, that it allows

enormous simplification of the unity rank performance index at the onset of the problem, with attendant savings in numerical quality and quantity. This result will also find application in the next chapter in another context.

A special case of Theorem 3.4 reveals an interesting structure which is well worth recording.

Corollary 3.4

If matrix Q_1 of Theorem 3.4 is such that its factor h forms a polynomial $h^T S(s)$ which has all zeros with non-positive real parts; then the factor H_2 of matrix Q_2 constructed as described in the proof will form two polynomials,

$$\begin{bmatrix} F_1(s) \\ F_2(s) \end{bmatrix} = H_2^T S(s),$$

having simple zeros (except possibly at the origin) which are restricted to lie on the imaginary axis where they occur in conjugate pairs and alternate with each other.

Proof

The proof only requires that the root location property described be demonstrated. The polynomials of interest are easily recognized as

$$F_1(s) = \mathrm{Ev}(F)$$

$$F_2(s) = \mathrm{Od}(F)$$

$$\text{and} \quad F(s) = h^T S(s).$$

It is a well-known theorem of Network Theory that the Ev(F) and Od(F) functions of a Hurwitz polynomial, F, have the root partitioning property recounted in the corollary [G6]. This phenomenon is referred to as the Alternation [G6] or Separation [T3] property.

A complex plane diagram of the respective zero locations of a typical case is useful to visualize this result.

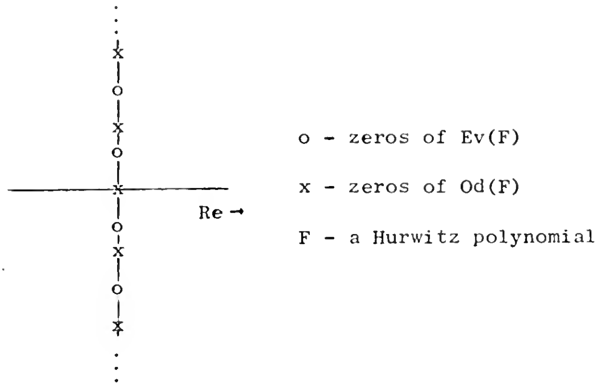


Figure 3.5 Zero Locations of Ev(F) and Od(F)

This corollary implies more than appears in the hypothesis. By Theorem 3.2 a unity rank Q, with $h^T S(s)$ a Hurwitz polynomial, can always be constructed if $\Psi(\omega)$ is non-negative, then the sparse matrix of the corollary can always be constructed as well.

3.6 Résumé of Ψ-Invariant Matrices

Because of the necessity of investigating observability and the testing for the absence of conjugate points much of the heuristic appeal of the form of matrices invariant under the operator $\Psi(Q;\omega)$ is obscured. In order to distinguish between matrices equivalent in the sense of Theorem 3.3 and matrices which result in identical Ψ's, without resorting to cumbersome phraseology, the latter will be referred to as "Ψ-invariant matrices." The conjugate point condition

is satisfied a priori for a very large class of weighting matrices (i.e., positive semidefinite) and the observability restriction has been shown to be a relatively innocuous constraint between Ψ-invariant matrices. Then a great deal is to be gained from a review of the structure of invariant matrices.

General Structure

Corollary 3.3 reveals that two matrices are Ψ-invariant if and only if their elements are such that the n-tuple

$$p_i = q_{ii} - 2q_{i-1,i+1} + 2q_{i-2,i+2} - \cdots \qquad i = 1,2,\ldots,n$$

$$(q_{ij} = 0 \quad \text{for} \quad i < 1 \quad \text{or} \quad j > n)$$

is equal for both matrices. For instance, in the third order case this can be interpreted as meaning that Ψ will remain invariant if a quantity is added to $q_{1,3}$ and $q_{3,1}$ and twice that quantity is added to q_{22}. In general, matrices which are Ψ-invariants of a given matrix may be constructed by manipulating the elements on "diagonals" running from lower left to upper right. This can be represented pictorially as in Figure 3.6.

$$\begin{bmatrix} 1 & x & 2 & x & 3 & \\ x & 2 & x & 3 & & \\ 2 & x & 3 & & & x \\ x & 3 & & \cdots & x & n-1 \\ 3 & & & x & n-1 & x \\ & & x & n-1 & x & n \end{bmatrix}$$

Figure 3.6 Structure of Ψ-Invariant Matrices

The integer entries of the illustration indicate which term of the n-tuple is affected by the element of a Q matrix which would be in that location; the x's are, of course, entries which influence no term. From this it is easy to see that matrices Ψ-invariant to a given matrix can be identified by inspection.

Spectral Factorization

A special case which is of considerable interest is the Q formed from the coefficients of the Hurwitz spectral factor of $\Psi(\omega)$. It is a form which can always be constructed when $\Psi(\omega)$ is non-negative and always meets the observability and conjugate point criteria. When the $i+j$ odd terms are zeroed the resulting matrix is always equivalent and possesses the root partitioning property of Corollary 3.4. This form is theoretically straightforward to compute but in practice is among the most difficult.

Diagonal

Another special case which is appealing in its simplicity is the diagonal Q. This matrix is formed by placing the coefficients of $\Psi(\omega^2)$ along the diagonal with the constant term first. A Ψ-invariant diagonal matrix also can always be formed; no observability difficulties will be encountered if the terms are all non-zero and the conjugate point condition will be satisfied if the terms are non-negative.

A simple example helps to demonstrate how this all fits together.

Example 3.3

Consider the real, even polynomial,

$$\Gamma(\omega) = \omega^4 + 2\omega^2 + 1,$$

and some members of the set of (Ψ-invariant) matrices,

$$\{Q \mid \Psi(Q;\omega) = \Gamma(\omega)\} .$$

i) Diagonal

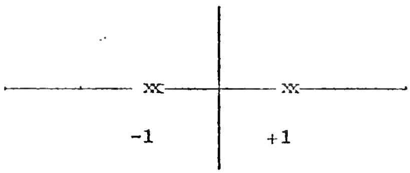
The diagonal Ψ-invariant matrix is simply,

$$Q_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which is non-singular (thus observable) and also positive definite.

ii) Unity Rank

The roots of $\Gamma(\omega)$ are sketched in the complex plane diagram below.



If the left half-plane zeros are chosen, the resulting unity rank $Q = HH^T$ is

$$Q_2 = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} [1 \quad 2 \quad 1]$$

and zeroing the $i+j$ odd terms results in

$$Q_3 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix} .$$

This matrix is clearly rank two and can easily be shown a Ψ-invariant of Q_1 by adding -1 to q_{13} and q_{31} and twice that (-2) to q_{22}. The zeros of $H^T S(s)$ for Q_3 are, respectively, $\pm j$ for the first output and 0 for the second, demonstrating the separation property of Corollary 3.4. If the factor is composed of one zero from each half-plane, the resulting Q is

$$Q_4 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} ,$$

which is again clearly an invariant of Q_1.

Note that in every Q matrix gener ed in this example the q_{11} and q_{nn} entries remain fixed. From Corollary 3.3 it is obvious that this must always be the case and that these terms are, respectively, the first (constant term) and last coefficients of $\Psi(\omega)$. The constancy of these elements has a very interesting physical interpretation which will be discussed in the next chapter.

3.7 Summary

The introduction to this chapter alluded to a concern which has been often expressed, with varying degrees of vehemence, by those well-versed in classical design techniques; that is, the minimality of some

performance measure is seldom directly relevant to a practical design problem. For the quadratic performance index and a linear system, however, it was shown that resulting optimal system configurations possess characteristics which are very much in the spirit of good system design. The minimum optimal system stability margins of 60° in phase and 6 db in gain compare favorably with those encountered in practice.

The problem of actually constructing a performance index which is minimized by an optimal system was found to lead, with due consideration for certain cancellation and existence difficulties, to an equivalence class of weighting matrices which enjoy some extremely enlightening and useful mutual properties. When a single member of an equivalence class has been determined, the remainder of the class can be constructed with relative ease.

CHAPTER IV

THE INVERSE PROBLEM AND LINEAR REGULATOR DESIGN

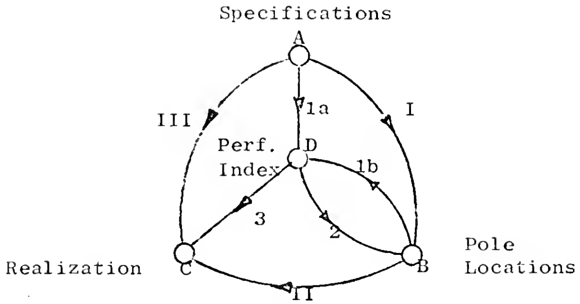
4.1 Introduction

The previous chapters have dealt with optimal systems, their properties and construction. It is essential that these concepts be well in hand if the techniques of optimal control theory are to be applied to design problems where a relevant performance measure is not readily discernible. In general, this is the case when a linear regulator is to be designed using optimal control theory. This chapter will investigate how optimal control theory and classical techniques may be used to complement one another in practical design problems.

For the purposes of this chapter it is helpful to think of classical synthesis techniques for linear control systems in terms of Figure 4.1. The problem originates with a linear system and a number of performance specifications that the compensated system is to satisfy (node A). The desired result is a realization (node C) which meets the performance criteria and contains a compensator which is satisfactory from the standpoint of practicality constraints, such as realizability and noisy or incomplete measurements.

One way of approaching the problem is to prescribe pole locations which will meet the performance specifications (node B) and then to construct a compensator which will place the plant poles in approximately the desired locations while not contributing significantly

to system response (path II). This is a quite prevalent philosophy, although it is often obscured by the specific design procedure used. Path I can be thought of as classical synthesis and analysis procedures used iteratively to arrive at these pole locations.



Classical Design Procedure		Alternatives	
I.	Prescribe pole locations	1.	Compute performance index
			a. from specifications b. from pole locations
II.	Design compensator		
III.	Direct design	2.	Optimal control problem
		3.	Automated optimal compensator design

Figure 4.1 Control System Design Procedures

A second approach is to determine at the onset the form of the compensator required and manipulate its parameters to arrive directly at a realization (path III). Well-known variations of this approach are Evans' root locus and lead-lag design [E2]. These techniques have the decided advantage of defining a clear-cut way to proceed but it may be difficult to accommodate noisy measurements.

The procedures of this chapter provide alternative paths in the design scheme as illustrated in Figure 4.1 by the Arabic numbered

branches. This provides for much greater flexibility in how the problem is attacked and employs the power of optimal control theory to relieve some of the procedural or computational burden in some or all phases of the design.

In some cases it will be possible to define a performance index directly from classical specifications (path 1a) which will be minimized by a control law which satisfies the specifications. The specific control law can then easily be computed (path 2) or the corresponding compensators can be designed (path 3), using Kalman-Bucy filter theory [K8,K9] or one of the new techniques for automated compensator design [P1,P2].

As indicated by path 1b of the figure, it is possible to compute a performance index corresponding to a closed-loop (optimal) pole configuration and the remainder of the synthesis can be completed through the use of compensator design techniques discussed in the last paragraph.

Figure 4.1 does not reveal some of the additional flexibility allowed by these procedures. For instance, a performance index may be specified which meets only a subset of the specifications and may then be used as a basis for design iteration. The techniques also provide for the design of sampled-data compensators (Section 4.4).

4.2 Pole Placement by Performance Index Designation

If a set of closed-loop system poles are proposed as a preliminary or final design configuration, the prescription of a performance index which is minimized by this configuration (path 1b of Figure 4.1) may be of considerable value in the completion or refinement of the design. Figure 4.2 illustrates such a case.

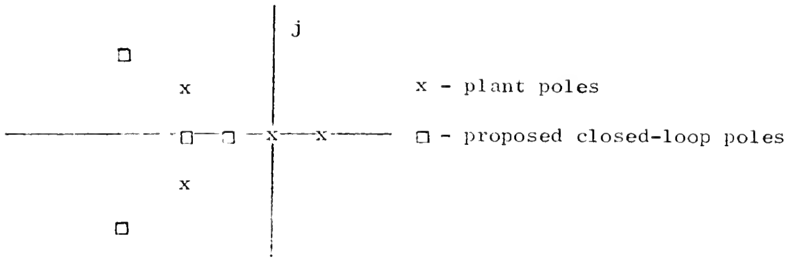


Figure 4.2 Proposed Closed-loop Pole Configuration

This configuration may have resulted from a root-locus type analysis, the exact scheme used is not germane to the present discussion; the important point is that it is not bound to any technique or form of compensation. The construction of a performance index which is minimized by this configuration (if possible) provides at most approximately $n^2/4$ nominal parameters[1] which may be varied to perturb the design or used to design the required compensator (path 3 of Figure 4.1).

The problem of computing a quadratic performance index which is minimized by specified closed-loop pole locations for a given plant is basically the problem entertained in great detail in the preceding chapter. This section seeks to deal with this problem on a more pragmatic basis.

The plant to be considered is a completely controllable, constant, linear system taken without loss of generality to be in companion matrix canonical form and the performance index is the infinite final time case of the last chapter.

[1]The precise number of pertinent parameters will be given later.

The stable closed-loop pole configuration of Figure 4.2 will be optimal with a positive semidefinite weighting matrix Q if and only if

$$\varphi(\omega) = |\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2 \geq 0 \qquad \text{for all real } \omega \tag{1}$$

where $\varphi(s)$ and $\varphi_k(s)$ are the normalized (highest coefficient is unity) open- and closed-loop characteristic polynomials. Although (1) is a succinct optimality criterion, it is not obvious how best to proceed to test a given polynomial (1) for non-negativity.

The necessity for testing the sign semidefiniteness of real even polynomials arises in many other applications, for instance, tests for positive reality in Network Theory [V1]. It can be shown by factoring $\Psi(\omega^2)$ or through the use of Sturm's Theorem that (1) is equivalent to requiring that $\Psi(\omega^2)$ have no positive real roots of odd multiplicity [V1, p. 106].

There is still not a computationally satisfactory approach evident. Calculation of the zeros of $\Psi(\omega^2)$, if the order is large, is difficult numerically and best avoided as a test. Recent results by Šiljak [S3] and Karmarkar [K10] extend the interpretation of the sign changes in the first column of the Routh table [G4] to test for positive real zeros of odd multiplicity. This test is probably the only alternative currently available to the computation of the zeros of $\Psi(\omega^2)$ as a necessary and sufficient test for non-negativity.

Appendix C outlines a numerical implementation (in Fortran IV) of a modification of Šiljak's method, which is both accurate and efficient; it is believed to be the only such program in existence.

It should be reiterated that condition (1) applies only to optimality with respect to a positive semidefinite Q; a system with $\Psi(\omega) \geq 0$ may well be optimal for a sign indefinite Q. In such a case the properties conditioned on Q being positive semidefinite (Section 3.3) would not in general hold but the Q equivalence relations and remarks concerning the observability requirements for Q (Chapter III)do.

If criterion (1) is satisfied, a positive semidefinite weighting matrix can be generated, in fact, an entire equivalence class of positive semidefinite Q's. In general the initial member of the equivalence class must be computed by spectral factorization. This operation, like the test for non-negativity, is not a numerically trivial one. The spectral factorization of $\Psi(\omega)$, if approached naïvely, consists of determining the 2(n-1) roots of Ψ, separating them by the sign of their real parts, constructing the factor composed of the zeros with negative real parts and multiplying it by the appropriate constant. This procedure is entirely unsatisfactory. It is well known that the confidence in approximation for root locations generally decreases drastically for polynomials of large order; this coupled with the error induced by constructing the spectral factor from its roots makes this procedure numerically hazardous.

An obvious alternative is to reduce the order of $\Psi(\omega)$ by substituting $\sigma = \omega^2$ and computing the roots of an $n - 1$ order polynomial which are the squares of the actual roots of interest. The roots of $\Psi(\omega)$ with negative real parts can be obtained directly. A refinement of this procedure is to compute quadratic factors of $\Psi(\sigma)$ only and from these factors, through some rather intricate logic, garner the

related left half-plane quadratic factors of $\Psi(\omega)$. This serves to reduce the total number of numerical manipulations, saving computing time and decreasing the sources of error propagation. The implementation of the spectral factorization algorithm discussed in the appendix takes this approach, utilizing an efficient technique due to Bairstow [K11], for the approximation of quadratic factors. Bairstow's iteration has been shown to have a rapid rate of convergence, although it is somewhat more sensitive to starting values than competing methods [K11, p. 101ff].

One objection to this procedure is that the roots of $\Psi(\omega)$ are not directly available as a secondary test for non-negativity. The test for non-negativity described earlier seems wholly adequate and any sacrifice of computational efficiency in the spectral factorization in favor of a redundant test does not appear justifiable.

The flow diagram of Figure 4.3 summarizes concisely the steps to be taken in the generation of a quadratic performance index which is minimized by a given plant and stable proposed (optimal) closed-loop pole configuration. The first step 1 is to compute the closed-loop and plant normalized characteristic polynomials (if not already known). The magnitude-square of the open- and closed-loop characteristic polynomials are next computed; this is another operation where the obvious procedure, i.e., multiplication of the respective polynomials by their complex conjugates, is not the best choice. Since the magnitude-square polynomials are even, half of the coefficients are zero. The technique outlined in Appendix C makes use of this structure by computing only the non-zero terms and in a way which obviates the use of complex

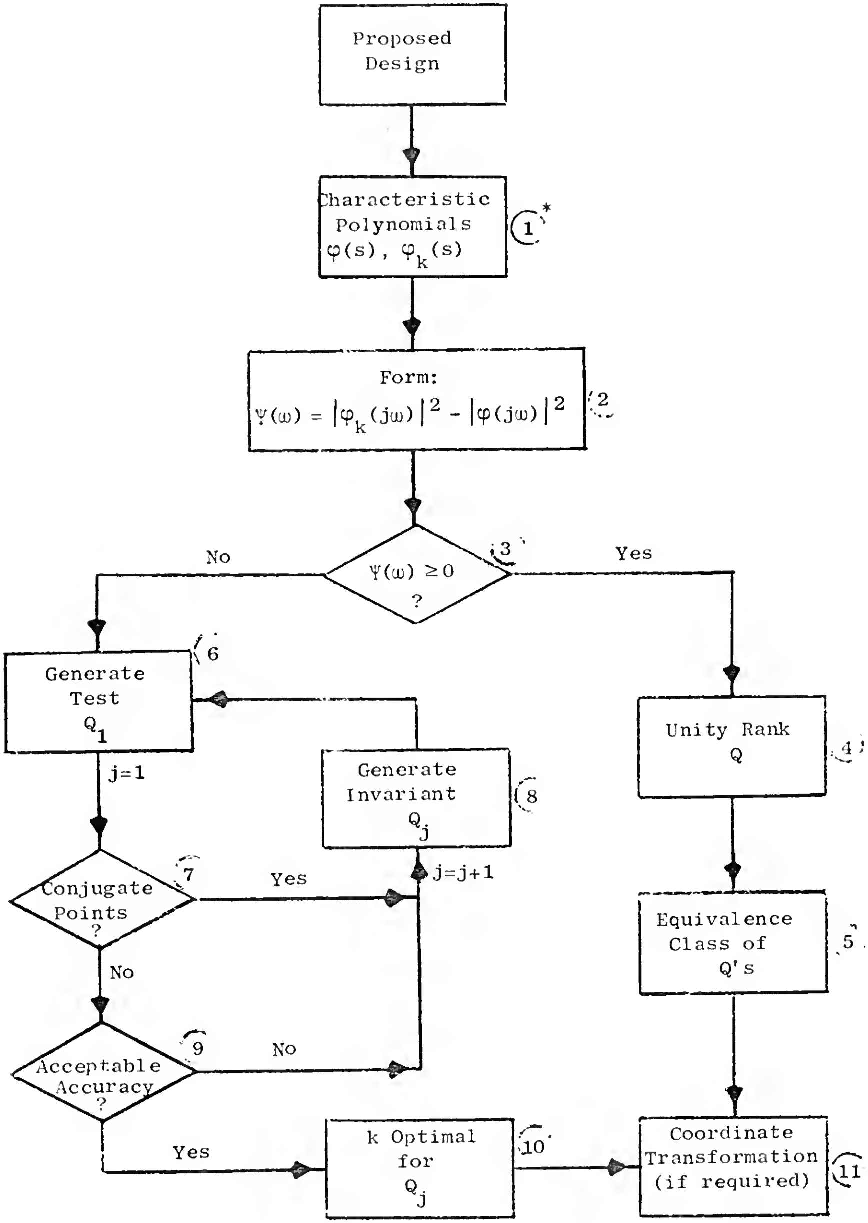


Figure 4.3 Procedure for Specification of a Performance Index from Proposed Pole Locations

---

*Numbers in circles are referred to in the text.

arithmetic. When the difference of the magnitude-square polynomials is taken (2), the resulting $\Psi(\omega)$ is an even polynomial of order $2(n-1)$ due to the normalization of the highest power coefficients of φ and φ_k to one.

A decision based on the non-negativity of Ψ is made in block (3). If Ψ is non-negative the system is clearly optimal and the specification of the Q's may proceed. The unity-rank Q is computed from the left half-plane factor of Ψ (4), using the techniques for spectral factorization discussed earlier, which further insures the observability conditions are satisfied, and the equivalence class of positive semidefinite Q's is immediately available (5).

If Ψ is not non-negative the procedure is no longer straightforward and positive results are not guaranteed. An initial Q_1 is constructed so that

$$\Psi(Q_1;\omega) = \Psi(\omega),$$

and all the unstable plant poles are observable (6). A reasonable choice for an initial Q is the diagonal case (coefficients of $\Psi(\omega^2)$ on diagonal); this is the easiest to generate and a desirable form for the answer. This choice of Q is inserted in the Riccati equation for the system in companion matrix canonical form,

$$\dot{P}(t) = -F^T P(t) - P(t)F + P(t)gg^T P(t) - Q, \quad P(t_f) = 0. \tag{2}$$

The Riccati equation is then integrated numerically until

$$\|\dot{P}(t)\| \leq \epsilon, \tag{3}$$

where ϵ is some prescribed non-negative bound or until a conjugate point intervenes (the solution becomes unbounded) (7). The non-linear

nature of the Riccati equation, which makes this sort of test necessary, also helps insure the success of the test, since it may be subject to the phenomenon of finite escape time [K2]. That is, as opposed to the solutions to linear differential equations which may be unbounded only in the limit, the Riccati equation may become unbounded in a finite time. Experience has shown that when a conjugate point does exist (2) generally becomes unbounded quite rapidly (in terms of the number of integration steps) and conversely when a conjugate point is not present steady-state (3) is reached promptly.

A third possibility is that (2) has neither a conjugate point nor a steady-state solution; that is, the solution is oscillatory. It is clear that if (2) is oscillatory it must also be periodic and hence the cost and the system response are periodic. This would contradict the supposition that the proposed system design was time-invariant and stable. Then, if a periodic solution to the Riccati equation occurs, it must be due to numerical errors and discarded in the same manner as a solution with a conjugate point.

A Q_j which results in a solution of (2) which diverges obviously must be discarded but it does not, in itself, indicate that the system is not optimal. The procedure branches to block 8 if a conjugate point is present and Q_{j+1} which is Ψ-invariant to Q_j is computed and the test for conjugate points 7 is repeated. It is difficult to conceive of an algorithmic scheme for "improving" Q_j in block 8. In fact, if this technique were known, one could apply it iteratively until the "best" Q is obtained and the subsequent test for conjugate points would be a general necessary and sufficient test for the

optimality of a control law An interactive strategy is ideal, Ψ-invariant Q's can be generated which not only are more likely to converge but also possess a desirable structure for the specific application.

The algorithm to be described seeks to generate an Ψ-invariant Q_{j+1} from a Q_j which fails the conjugate point condition 7 by shifting the negative eigenvalues of Q_j generally to the right while maintaining as much simplicity as possible in Q_{j+1}. This is done by testing the diagonal terms q_{ii} of Q_j for $1 < i < n$ progressively[2] until a negative entry is detected which is then deleted, using equation (17) of Chapter III.

Suppose that the minor shown below is a 3 x 3 principal minor, containing the negative diagonal entry b.

$$\begin{array}{cc|ccc|}
\hline
-d/2 & 0 & a & 0 & 0 \\
0 & 0 & 0 & b & 0 \\
0 & 0 & 0 & 0 & c \\
\hline
\end{array}$$

Because the first Q (Q_1) was diagonal and since the method proceeds progressively down the diagonal, the remainder of the rows shown are null except the first which may have an off-diagonal entry -d/2 placed by the preceding iteration. By the results of the last chapter, altering this minor as shown below:

---

[2] As noted in the preceding chapter, the first and last terms on the diagonal cannot be altered and invariance maintained.

$$\begin{array}{cc|ccc|}
\hline
-d/2 & 0 & a & 0 & -b/2 \\
 & & 0 & 0 & 0 \\
 & & -b/2 & 0 & c \\
\hline
\end{array}$$

leaves the Ψ of the resulting matrix invariant. Each time a Q_j fails the conjugate point test the described procedure removes an additional negative diagonal entry until all have been eliminated. In this manner the scheme allows k+1 iterations where k is the number of negative coefficients of $\Psi(\omega^2)$ (or the number of negative diagonal terms) before the system is discarded. The procedure is executed iteratively, rather than eliminating all of the negative terms in one step, with the hope of determining as simple (diagonal) a Q as possible for which the system is optimal.

At this point the justification for such a procedure is probably not clear The object of the algorithm is to generate a sequence of invariant Q's whose positive diagonal elements dominate their respective rows. This will hopefully result in a shifting of the eigenvalues to the right. Central to this argument is the Gerschgorin Circle Theorem which was recorded as Lemma 3.5 in the previous chapter.

Since the rows of any Q_j generated by the described technique have at most two off-diagonal elements, each being half of a negative diagonal term, the eigenvalues of the modified matrices will be increasingly restricted to be "close" to the right half-plane if the

diagonal elements are approximately of the same magnitude.[3] There is no hope that the matrices will eventually become positive semidefinite with repeated iterations (this would require $\Psi(\omega) \geq 0$ by Theorem 3.2); however, it is reasonable to expect the likelihood of passing the conjugate point condition to increase if the eigenvalues are shifted to the right. An example will help to clarify the details.

Example 4.1

Consider the sequence of Q's which would be generated by this method for

$$\Psi(\omega) = 7\omega^{12} - 6\omega^{10} - 5\omega^{8} + 3\omega^{6} + 2\omega^{4} - 4\omega^{2} + 1.$$

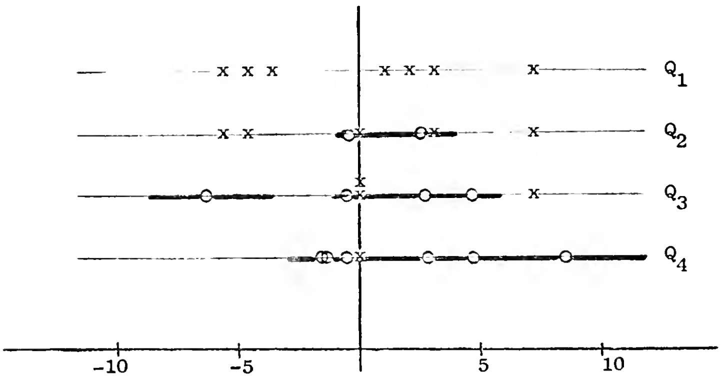
This Ψ is clearly not non-negative, e.g., $\Psi(1) = -2$; then there is no possibility of constructing a positive semidefinite Q for it. The procedure beginning with block 6 of the flow chart of Figure 4.3 must be called upon. Since there are three negative coefficients of Ψ there will be at most four Ψ-invariant Q's generated; Figure 4.4 illustrates these iterations. The first, Q_1, obviously has eigenvalues which are simply the coefficients of $\Psi(\omega^2)$. The first iteration removes the first negative entry on the diagonal (-4) by "splitting" it into the two off-diagonal terms resulting in Q_2. Four of the eigenvalues remain unchanged, the one that was at -4 is moved to the origin and the remaining two have been "smeared" by the interpretation given by Gerschgorin's Theorem. The eigenvalue which was at 1 may now be anywhere within the interval (-1,+3) and the eigenvalue at 2 may now lie in the region (0,+4) as shown in the plot.

[3] Actually the right half-line, since the matrices under consideration are symmetric and consequently have only real eigenvalues.

$$\Psi(\omega) = 7\omega^{12} - 6\omega^{10} - 5\omega^{8} + 3\omega^{6} + 2\omega^{4} - 4\omega^{2} + 1$$

$$Q_1 = \begin{bmatrix} 1 & & & & & & \\ & -4 & & & \bigcirc & & \\ & & 2 & & & & \\ & & & 3 & & & \\ & & & & -5 & & \\ & \bigcirc & & & & -6 & \\ & & & & & & 7 \end{bmatrix} \qquad Q_2 = \begin{bmatrix} 1 & 0 & 2 & & & & \\ 0 & 0 & 0 & & \bigcirc & & \\ 2 & 0 & 2 & & & & \\ & & & 3 & & & \\ & & & & -5 & & \\ & \bigcirc & & & & -6 & \\ & & & & & & 7 \end{bmatrix}$$

$$Q_3 = \begin{bmatrix} 1 & 0 & 2 & & & & \\ 0 & 0 & 0 & & \bigcirc & & \\ 2 & 0 & 2 & & & & \\ & & & 3 & 0 & 2.5 & \\ & & & 0 & 0 & 0 & \\ & \bigcirc & & 2.5 & 0 & -6 & \\ & & & & & & 7 \end{bmatrix} \qquad Q_4 = \begin{bmatrix} 1 & 0 & 2 & & & & \\ 0 & 0 & 0 & & \bigcirc & & \\ 2 & 0 & 1 & & & & \\ & & & 3 & 0 & 2.5 & \\ & & & 0 & 0 & 0 & 3 \\ & \bigcirc & & 2.5 & 0 & 0 & 0 \\ & & & & 3 & 0 & 7 \end{bmatrix}$$



o - Computed eigenvalues
x - Eigenvalues by inspection
▬ - Permitted eigenvalue regions

Eigenvalue Distributions

Figure 4.4 Eigenvalue Distributions for Q_j

With subsequent iterations the original eigenvalues are spread into permitted regions which are generally to the right. Due to the structure of such matrices the actual eigenvalues are easily computed and included in the plot. The process lives up to expectation, the negative eigenvalues are moved distinctly to the right and the likelihood of passing the conjugate point condition (esp. for Q_4) is considerably enhanced.

Note that the rank of Q_j for $j > 1$ is reduced by at least one. This causes concern that some observability may be lost in the process, with the result that an unstable plant pole becomes unobservable (a stable plant pole becoming unobservable is, of course, inconsequential). Although this may indeed be the case, it will be shown later that sufficient safeguards exist to prevent it from passing unnoticed.

Within the framework of the process outlined there exist many variations which may be of value in certain applications. For instance, it may be advantageous to dispatch the negative diagonal terms of largest modulus first rather than approaching them progressively along the diagonal. This modification applied to Example 4.1 leads to a somewhat greater "improvement" in the second iteration than the method originally employed, but the results of the final iteration are, of course, identical. The important observation to make is that no single variation will be best in every case and that employment of good judgment at this point in the scheme, rather than following a strictly algorithmic approach, will probably be rewarded with simpler resulting Q's. It is for this reason that the suggestion was made earlier to utilize an interactive strategy in a computer implementation of this procedure.

If the iterative technique of modifying Q_j and testing for conjugate points succeeds in obtaining a Q for which the Riccati equation does not diverge, the next step is to determine if this Q is actually a solution to the inverse problem. This is necessary to guard against gross numerical errors in the integration of the Riccati equation falsely indicating the absence of conjugate points, and, as mentioned earlier, to insure that a detrimental loss of observability from Q has not occurred.

The steady-state solution, P, to the Riccati equation (2) is available from the conjugate point test (block 6) and provides this check. The feedback control law,

$$k = Pg,$$

is simply the last column (and row) of $\bar{P}$. By the structure of the companion matrix canonical form if

$$\varphi(s) = s^n + a_n s^{n-1} + \ldots + a_2 s + a_1,$$

then $$\varphi_k(s) = s^n + (a_n+k_n)s^{n-1} + \ldots + (a_2+k_2)s + (a_1+k_1),$$

where k_i is the ith entry of k. The coefficients of φ_k computed by substituting $p_{i,n}$ for k_i can be compared to the correct (original) values. This test is represented by block 10 of Figure 4.3.

In passing, it should be noted that the process for generating successive Q's outlined here fails to make full use of the relation of the elements of invariant matrices (Corollary 3.3) in that only the terms on the diagonal and second superior diagonal are affected (i.e., q_{ii}, $q_{i+2,i-2}$ and $q_{i-2,i+2}$ when $q_{ii} < 0$). This is because the process becomes considerably more difficult to justify as the number of altered

terms increases and because it was felt that the additional complexity in the resulting Q was self-defeating. However, further research in this area may ultimately lead to a necessary and sufficient condition for optimality when Q is not positive semidefinite.

There remains but one operation which may be required to complete the process of designing a performance index for which a given closed-loop pole configuration is optimal. If the state model for the plant (assumed completely controllable) is not in companion matrix canonical form it may be desired to transform the Q which has been generated into the coordinate system of the plant. This is accomplished by the congruent transformation,

$$\tilde{Q} = T^T Q T,$$

where $\tilde{Q}$ is in the original coordinate system of the plant and T is the non-singular matrix of the similarity transformation (Section 3.4),

$$\hat{F} = T\tilde{F}T^{-1},$$

which transforms the original state matrix ($\tilde{F}$) into a companion matrix ($\hat{F}$).

It is necessary to show that this can be done without loss of generality. That is, that the closed-loop poles are invariant and that no conjugate points are introduced under this transformation.

Theorem 4.1

> Optimality is preserved under similarity transformation. That is, if a completely controllable linear plant and feedback control law are optimal with respect to a quadratic performance index, the system and control law subject to a non-singular coordinate transformation minimizes the performance index in the new coordinate system.

This theorem can be easily proven by applying the transformation required to return the system to its original form from the companion matrix formulation to the Riccati equation. Then, by the non-singularity of T, if the Riccati equation becomes unbounded in one coordinate it will in the other as well.

4.3 Design by Performance Index Iteration

Once a technique, such as the one presented in the last section, is available for generation of performance indices for optimal closed-loop pole configurations a simple design scheme becomes evident. The performance index could be used as a basis for improving a first attempt at design. In terms of Figure 4.1, this would entail traversing branches 1b and 2 iteratively, each time altering the entries of Q in a manner intended to meet additional members of the specification set.

This approach offers several advantages over competitive classical design techniques. If the preliminary closed-loop pole configuration is chosen so that

$$\Psi(\omega) = |\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2 \geq 0 \qquad \text{for all real } \omega,$$

then maintenance of the positive semidefiniteness of the resulting Q throughout all the iterations will insure that the important stability criteria of gain and phase margin are at least 6 db and 60°, respectively.

This permits the designer to concentrate on bringing transient response to within desired limits and later returning to the stability margins if more stringent ones are required.

A second advantage is that a tractable number of design parameters are displayed in the Q matrix. The precise number can easily be shown to be $\left[\frac{(n+1)^2}{4}\right]$, where the heavy brackets represent the greatest integer function; i.e.,

$$[a] = \text{greatest integer} \leq a.$$

This quantity does not include the duplicate (by symmetry) terms appearing across the diagonal of Q.

This means that for a third-order plant there are at most four parameters of Q which are pertinent to the design; similarly for a tenth order plant, at most 32. These numbers may seem rather large in relation to the number of coefficients of the characteristic polynomials; however, these parameters, as will be seen later, may be related more or less directly to the transient response of the closed-loop system, in contrast to the classical design parameters. Further, far less than the maximum number given will be generally required; in fact, usually it will only be necessary to vary two of them [H2].

It was noted in the last chapter that the first and last terms on the diagonal of equivalent Q matrices must remain invariant. That is, q_{11} must be the same among all members of an equivalence class and similarly q_{nn} must not change. The next theorem presents an important new result which is a consequence of this observation.

Theorem 4.2

If Q is any member of an equivalence class of weighting matrices (Definition 3.1) for a scalar nth order linear system and optimal control law in companion matrix canonical form, then

$$q_{11} = \prod_{i=1}^{n} R_i^2 - \prod_{i=1}^{n} r_i^2$$

and

$$q_{nn} = \sum_{i=1}^{n} R_i^2 - \sum_{i=1}^{n} r_i^2 ,$$

where r_i and R_i, $i = 1,2,3,\ldots,n$

are the plant and optimal closed-loop poles, respectively.

Proof

In the discussion of the diagonal Q case in Section 3.6, it was shown that the coefficients of $\Psi(\omega^2)$ form the diagonal entries. The optimization problem implied by this Q may have conjugate points. Nonetheless, since the control law of the hypothesis was specified to be optimal, there must exist at least one weighting matrix which forms a performance index minimized by the control law and further it must be a Ψ-invariant of the diagonal Q. Then, since the first and last entries on the diagonal of invariant matrices cannot change, these entries must be exactly the first and last coefficients of $\Psi(\omega^2)$.

Consider the monic polynomial with real coefficients,

$$p(s) = s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_o, \qquad (4)$$

and its magnitude square

$$|p(j\omega)|^2 = \omega^{2n} + b_{n-1}\omega^{2(n-1)} + \ldots + b_1\omega^2 + b_o. \qquad (5)$$

By definition

$$\Psi(\omega) = |\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2; \tag{6}$$

that is, Ψ is the difference of two polynomials similar to (5). Then the constant term of Ψ, q_{11}, is the difference of two terms corresponding to b_o in (5). Clearly,

$$b_o = a_o^2,$$

and by a well-known theorem of elementary algebra, the constant term of a monic polynomial is (to within the sign) the product of its zeros. Then b_o is the square of the product of the roots of (4) and application of this abstraction to (6) demonstrates the expression for q_{11}.

The coefficient of $\omega^{2(n-1)}$ of Ψ, $q_{n,n}$, is the difference of terms similar to b_{n-1} in (5). By examining the product of $p(j\omega)$ and its complex conjugate it is easily verified that

$$b_{n-1} = a_{n-1}^2 - 2a_{n-2}. \tag{7}$$

Again it is known from elementary algebra that for a monic polynomial such as (4), a_{n-1} is the negative of the sum of the zeros of (4) and a_{n-2} is the sum of all combinations[4] of products of the roots of (4) taken two at a time. That is,

---

[4] Here combination is used in the sense of combinatorial analysis; i.e., no distinction is made between the ordered pairs (c,d) and (d,c).

$$a_{n-1} = -\sum_{i=1}^{n} r_i$$

and

$$a_{n-2} = \sum_{i=2}^{n} \sum_{j=1}^{i-1} r_i r_j,$$ where $r_i, i=1,2,3,\ldots,n$ are the roots of (4). The expression for a_{n-2} can be rewritten by forming the sum of all the possible permutations of products of two roots (including with themselves) and subtracting off the sum of the roots squared and dividing by two to account for the combinations being summed twice; i.e.,

$$a_{n-2} = \frac{1}{2}\left[\left(\sum_{i=1}^{n} r_i\right)^2 - \sum_{i=1}^{n} r_i^2\right].$$

Then, using this expression for a_{n-2} and substituting into (7) results in

$$b_{n-1} = \sum_{i=1}^{n} r_i^2 .$$

The expression for q_{nn} then follows directly from application of this result to equation (6).

The interpretation of the effect of q_{nn} on a system can be greatly enhanced through the use of a simple mechanical analogy. If the poles (r_i) of a linear system are thought of as unit point masses, the expression,

$$I = \sum_{i=1}^{n} r_i^2,$$

would be the resulting moment of inertia with respect to a perpendicular axis of rotation through the origin such a mechanical system would possess. Then q_{nn} represents the amount by which the "moment

of inertia" of the system has increased with the addition of optimal feedback.

If q_{nn} is non-negative the optimal closed-loop system will have at least one pole further removed from the origin than the plant poles. That is, for $q_{nn} > 0$ the closed-loop system will tend to be faster in response than the open-loop, and, in general, the larger q_{nn} is made the faster the optimal system.

Although q_{11} does not admit to analogy as well as q_{nn}, its effect on the optimal pole locations is similar to that of q_{nn}; that is, the speed of response also tends to increase with q_{11}.

Recently there has been a resurgence of experimental investigations of the effect of entries of the Q matrix on the dynamic response of the closed-loop system [e.g., H2,R1]. Although it is hopeless to attempt to glean a generalized design scheme from such studies, since they are bound to specific plant configurations (i.e., a fixed number of poles and zeros), it is possible to obtain some guidelines. Houpis and Constantinides [H2] observed that q_{11} and q_{nn} together have primary effect on rise time (t_r) and setting time (t_s) and the remainder of the entries primarily influence overshoot. These observations seem to coincide quite well with the analysis given here.

At this point a design example would help to solidify the preceding remarks.

<u>Example 4.2</u>

It is desired to compensate the unstable plant

$$u \longrightarrow \boxed{\frac{1.0}{s(s-1)(s+2)}} \longrightarrow y$$

so that the following performance specifications are met:

1. Gain Margin: $\overline{GM} \geq 12$ db
2. Phase Margin: $\varphi M \geq 60°$

and in response to a step input,

3. Overshoot: $Os \leq 5\%$
4. Rise time (within 90%): $t_r \leq 1$ sec
5. Setting time (within 1%): $t_s \leq 2$ sec.

The approach taken here will be to develop a preliminary design, using dominant roots techniques, which is also optimal, and to then iterate on the entires of the Q matrix until a suitable final design is arrived at. By dealing only with positive semidefinite Q's, phase margin, and probably gain margin as well, will be guaranteed, leaving only the dynamic response specifications to be investigated.

Using the dominant root philosophy, the complex poles should have a time constant of approximately 2 seconds to meet the requirement for t_s and a damping of approximately 0.7 in order to keep maximum overshoot within 5% and still meet rise time specifications [D4, Fig. 4-3, p. 91]. The third pole removed by 2.5 times the dominant time constant should insure dominance. The resulting preliminary design pole configuration is shown in Figure 4.5.

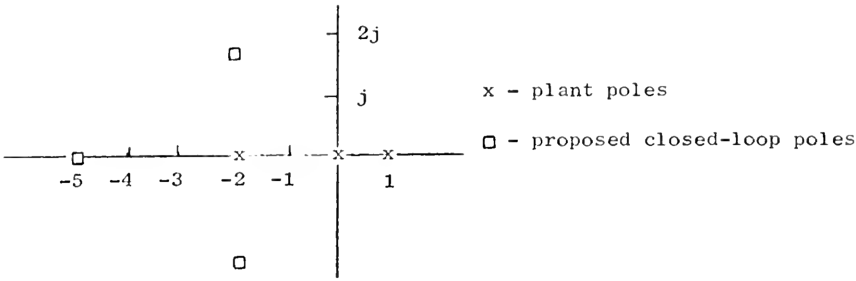


Figure 4.5 Preliminary Design Pole Configuration

The algorithm of the preceding section is now applied to obtain a Q matrix (if one exists). The system is indeed optimal and a Q matrix which will work is

$$Q = \begin{bmatrix} 1600 & 0 & 0 \\ 0 & 60 & 0 \\ 0 & 0 & 20 \end{bmatrix} .$$

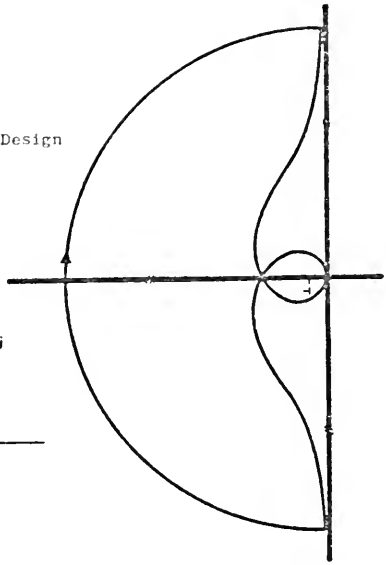
The transient response of this system was computed, using the quadratic optimization and simulation program LQL [B7,B9]. The important characteristics of the response are recorded in the first row of the table of Figure 4.6. It is clear that response is too slow, hence, q_{11} is increased; however, continued increase much beyond the value of entry 3 causes the overshoot specification to be exceeded. Reducing q_{33} has the effect of reducing the damping on the dominant poles (decreasing t_r) and bringing the third pole in toward the origin and results in more rapid settling (decreasing t_s). Run number 6 meets the specifications on transient response.

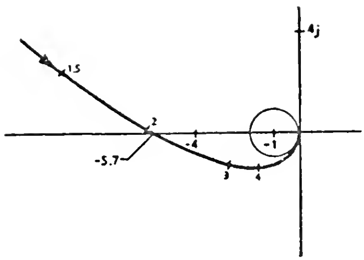
Run No.	q_{11}	q_{22}	q_{33}	t_r (sec)	Os (%)	t_s (sec)
1	1600	60	20	1.2	3.3	2.6
2	576	60	20	1.6	2.0	3.2
3	3000	60	20	1.1	4.0	2.3
4	2500	60	10	1.1	4.1	2.2
5	2500	60	4	1.1	4.0	2.1
6	3000	60	4	1.0	4.6	2.0
7	5000	60	4	0.9	5.3	1.8

Figure 4.6 Transient Response vs q_{ii}

The stability margins only remain to be investigated. Since the Q for all the iterations is positive semidefinite, GM ≥ 6 db and φM ≥ 60°. The Nyquist diagram of Figure 4.7 shows that the gain margin is considerably in excess of 12 db; thus iterative number 6 meets all the design specifications.

Figure 4.7 Nyquist Plot of Final Design





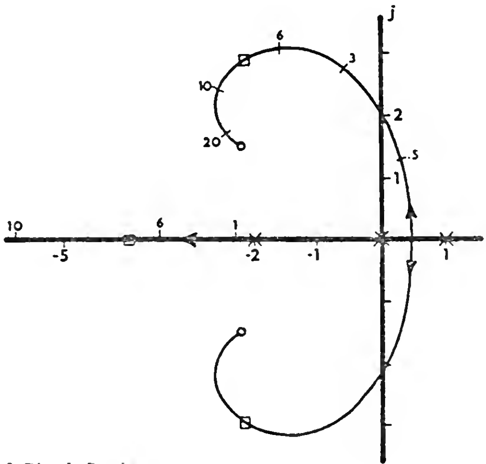


Figure 4.8 Root-Locus Plot of Final Design

× - Plant poles

□ - Closed-loop poles

o - Feedback zeros

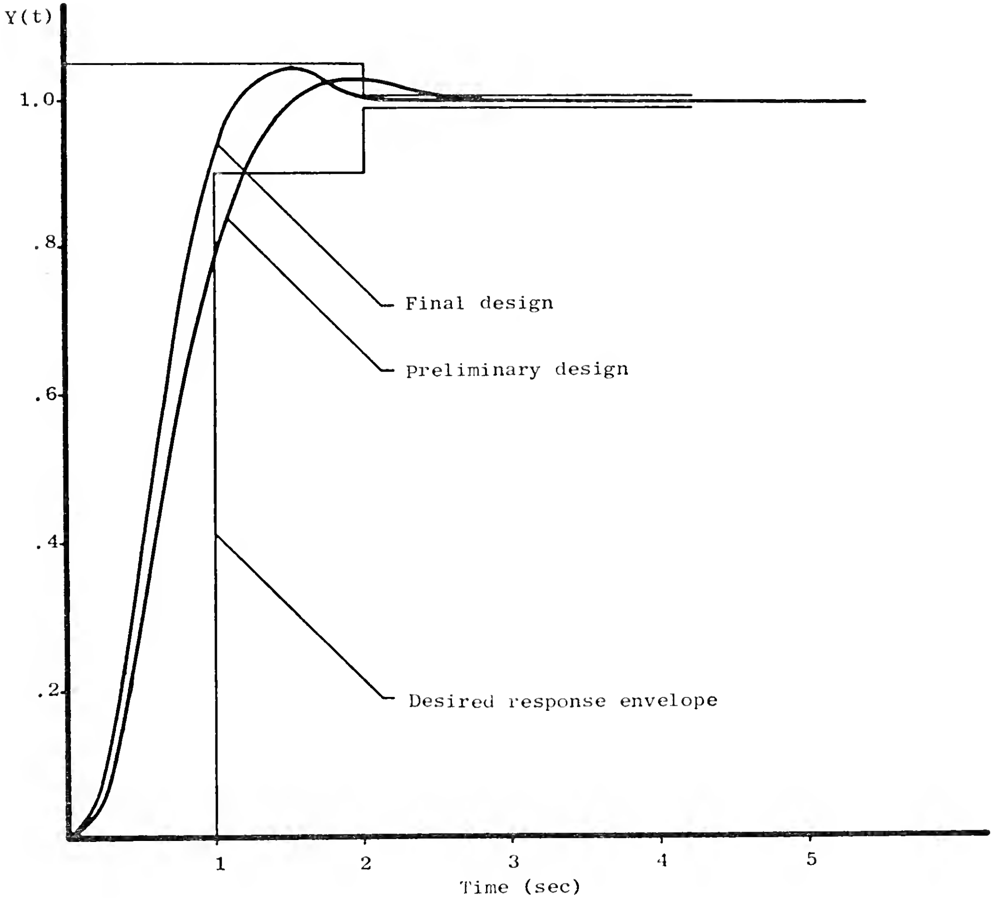


Figure 4.9 System Responses to a Step Input - Preliminary and Final Designs

The final design meets the performance criteria:

$$\overline{GM} = 15.18 \text{ db}$$
$$\varphi M = 62.9^\circ$$
$$t_r = 1.0 \text{ sec}$$
$$t_s = 2.0 \text{ sec}$$
$$Os = 4.6\%$$

and has the state feedback control law

$$k = \begin{bmatrix} 54.77 \\ 33.53 \\ 7.49 \end{bmatrix} .$$

Although Figure 4.6 only tabulates seven iterations on the Q matrix, fourteen iterations using program LQL were actually made, requiring approximately 6 seconds of CPU time on an IBM 360/65 at a cost of about $0.65.

This example illustrates two facets of this method. First, there are no concrete rules, as such, for the manipulation of the entries of Q to obtain desired changes in response, only general guidelines. Secondly, the speed and low cost with which the iterations can be computed make this process of practical value.

A final feature of this scheme is that the resulting design is not bound to any specific form of compensation: this can be regarded at times either as an advantage or a handicap However, with the rapid advance being made in automated compensator design [e.g., P1, P2], the disadvantages associated with the compensator not being prespecified are becoming largely illusionary.

4.4 Design by Explicit Performance Index Specification

If the design problem is more fully specified, a useful variation of the preceding scheme will often be profitable. For instance, when specifications of the sort required for Example 4.2 are given and in addition it is desired to minimize (in a mean-square sense) a specific system output, or a weighted sum of outputs, the technique of performance index iteration takes on a particularly interesting form.

Consider the scalar input, multioutput linear system,

$$\dot{x} = Fx + gu$$

$$y = H^T x,$$

for which it is desired to determine a feedback control law so that a given (but at the moment arbitrary) set of classical performance specifications are satisfied and in addition the performance index,

$$J = \int_0^\infty (cy^T Ay + u^2)dt, \qquad (8)$$

is minimized, where c is an unspecified scalar and A a symmetric weighting matrix. Then in the usual notation,

$$Q = cHAH^T$$

and $$c\Psi(\omega) = c\Psi(HAH^T;\omega),$$

which can be easily computed using Lemma 3.4. Any closed-loop system which minimizes (8) will have a characteristic polynomial, $\varphi_k(s)$, such that

$$|\varphi_k(j\omega)|^2 = |\varphi(j\omega)|^2 + c\Psi(\omega) \qquad (9)$$

by Theorem 3.3. Then if the substitutions

$$p_k(\sigma) = |\varphi_k(j\omega)|^2 \Big|_{\sigma=\omega^2},$$

$$p(\sigma) = |\varphi(j\omega)|^2 \Big|_{\sigma=\omega^2},$$

and

$$\Psi(\sigma) = \Psi(\omega) \Big|_{\sigma=\omega^2}$$

are made into (9), the resulting expression,

$$p_k(\sigma) = p(\sigma) + c\Psi(\sigma),$$

is reminiscent of the classical root-locus form, except that the poles of interest are not those located on the locus but rather their (negative) square roots. Such plots have been called "Root-Square-Locus" plots by Chang [C3] who made use of them in a different context to determine solutions to the so-called ISE problems.

Since the root-square-locus determines where the square of the closed-loop poles may lie to minimize (8), the parameter c is the only variable available for use in reconciling any remaining performance specifications. If the remaining specifications include bandwidth and steady-state error (for a constant input) the region of permissible values for c may be significantly reduced.

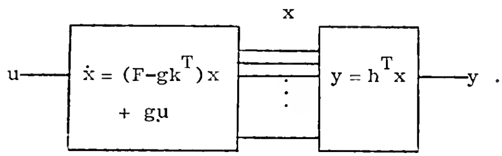
If A is positive semidefinite, then there exists a unity rank Q such that

$$Q = hh^T \sim H^T A H$$

and, further,

$$h^T S(s) = \psi(s),$$

is the Hurwitz spectral factor of $\Psi(\omega)$. Then the multioutput case can be reduced to an equivalent single output case if A is positive semi-definite. That is, for the purposes of the design, the system can be envisioned as:



When y is the equivalent single output derived above, the open-loop (k=0) and closed-loop transfer functions are, respectively:

$$G_o(j\omega) = \frac{\psi(j\omega)}{\varphi(j\omega)}$$

and

$$G_c(j\omega) = \frac{\psi(j\omega)}{\varphi_k(j\omega)} \quad .$$

It is reasonable to expect that any bandwidth or steady-state error specifications given are in terms of this equivalent output; in the single output case this would certainly be true. The multiple output case may require a redefinition of (8) to utilize the following procedure, depending on the interpretation given to bandwidth and steady-state error when more than one output is considered.

Then the bandwidth of the closed-loop system is ω_o such that,[5]

$$|G_c(j\omega_o)|^2 = \tfrac{1}{2}|G_c(0)|^2,$$

or equivalently,

$$\frac{\Psi(\omega_o)}{|\varphi_k(j\omega_o)|^2} = \frac{1}{2}\,\frac{\Psi(0)}{|\varphi_k(0)|^2} \quad .$$

---

[5] Clearly $G_c(s)$ must be low-pass, i.e., $\psi(s)$ has no zero at the origin, for this definition to be meaningful.

But $$|\varphi_k(j\omega)|^2 = |\varphi(j\omega)|^2 + c\Psi(\omega);\ \text{then}$$

$$\frac{\Psi(\omega_o)}{|\varphi(j\omega_o)|^2 + c\Psi(\omega_o)} = \frac{1}{2}\,\frac{\Psi(0)}{|\varphi(0)|^2 + c\Psi(0)}\ .$$

Denoting $\Psi(0)$ as p_1 and $\varphi(0)$ as a_1 (the coefficients of the least powers in Ψ and φ) and solving for c results in

$$c = \frac{|\varphi(j\omega_o)|^2}{\Psi(\omega_o)} - \frac{2a_1^2}{p_1}\ ,$$

$p_1 \neq 0$ (see footnote 5), which can be further simplified to

$$c = \frac{1}{|G_o(j\omega_o)|^2} - \frac{2a_1^2}{p_1}\ . \tag{10}$$

Then the c in (8) necessary for the optimal system to possess a specified bandwidth, ω_o, can be easily evaluated from expression (10) and is a function only of the frequency response of the plant.

If bandwidths much in excess of the bandwidth of the plant are required $|G_o(j\omega_o)|$ will become small, c will rapidly increase and the feedback gains necessary to accommodate the resulting large coefficients of $\varphi_k(s)$ will likewise increase. Clearly, if the amount of feedback gain is limited, the bandwidth of the optimal system cannot be much larger than that of the plant.

A similar procedure results in an expression for steady-state error for a unit step input. If steady-state error is defined as the difference between the closed-loop system input and output as $t \to \infty$, then using the final value theorem of transform theory, the final value of error (FVE) is simply

$$\mathrm{FVE} = 1 - G_c(0),$$

or

$$\mathrm{FVE} = 1 - \left(\frac{p_1}{a_1^2 + cp_1}\right)^{1/2} .$$

The square root is allowed and has the correct sign since $\psi(s)$ and $\varphi_k(s)$ are Hurwitz polynomials. Then, by simple manipulation, the c required for a specified FVE is

$$c = \frac{1}{(1-\mathrm{FVE})^2} - \frac{a_1^2}{p_1} , \tag{11}$$

for $p_1 \neq 0$ and $\mathrm{FVE} \neq 1$.

As should be expected, steady-state error and bandwidth are conflicting requirements. Whereas large bandwidths result in large values for c, increased accuracy follows from decreasing c. If zero FVE is specified, then from (11)

$$c = 1 - a_1^2/p_1 .$$

Substituting this value for c into (10)

$$\frac{1}{|G_o(j\omega_o)|^2} = 1 + \frac{a_1^2}{p_1}$$

or

$$|G_o(j\omega_o)|^2 = \frac{p_1}{p_1 + a_1^2} .$$

If the required bandwidth, ω_o, is greater than the open-loop bandwidth, then

$$|G_o(j\omega_o)|^2 < \frac{1}{2}\,\frac{p_1}{a_1^2}$$

and $$|G_o(j\omega_o)|^2 = \frac{p_1}{p_1 + a_1^2} < \frac{p_1}{2a_1^2} ,$$

hence, it is possible to obtain closed-loop bandwidth greater than that of the plant and zero FVE simultaneously if

$$p_1 > a_1^2 .$$

This condition can always be met by providing additional gain in the plant path. That is, if a_o is the constant coefficient of the numerator of $G_o(j\omega)$ and g is a variable gain on the output of the plant, then g is simply increased until

$$p_1 = (ga_o)^2 > a_1^2 .$$

Although the preceding treatment has dealt exclusively with a unity rank Q, the actual computation of the required optimal control law may, of course, be carried out using any Q equivalent to chh^T. These methods for computing optimal systems satisfying steady-state error and bandwidth specifications may also be profitably used as an initial design for the performance index iteration techniques (Section 4.3) when additional criteria are to be met.

In summary, when it is desired to remove the disturbances from a specific system output or weighted sum of outputs, additional techniques become available for selection of the correct performance index to meet design requirements. The squares of the permitted closed-loop locations may be identified by the use of the root-square-locus, and a performance index generated to place them anywhere along that locus. If steady-state error and/or bandwidth specifications are to be met,

performance indices may be simply constructed which result in optimal systems satisfying these criteria. In each case, these performance indices are explicitly defined.

4.5 Sampled-Data Controller Design

This section will consider a design problem which is particularly well adapted to solution by techniques arising from the inverse problem. The design of digital controllers for continuous systems is attacked, using methods which represent a new approach to a long-standing problem.

The specific problem to be dealt with is identical to the one considered earlier in this chapter with the exception that the resulting compensator is constrained to be a digital device rather than continuous. That is, it is desired to design a digital compensator for the linear, completely controllable, scalar input system

$$\dot{x} = Fx + gu \tag{12}$$

such that a given set of classical performance specifications are satisfied.

The approach will be to use the techniques of the preceding sections to design a performance index, having a corresponding continuous optimal control law which meets the performance specifications. The sampled-data model for the continuous system,

$$x_{m+1} = \Lambda x_m + \Gamma u_m, \tag{13}$$

is then computed so that the solutions to (12) and (13) agree at the sample points when the input is held constant over the sample interval,

as it would be if it were the output of a zero-order hold of a digital controller. The required matrices Λ and Γ in (13) can be determined [B7] from the solution to (12), i.e.,

$$x(t) = \Phi(t-t_o)x(t_o) + \int_{t_o}^{t} \Phi(t-\tau)gu(\tau)\, d\tau,$$

where $$\Phi(t) = e^{Ft},$$

by evaluating it over the interval, $(t_o = m\Delta T,\ t = (m+1)\Delta T)$, where ΔT is the sample period. Then with the convention that variables evaluated at time $m\Delta T$ be denoted by a subscript m,

$$\Lambda = \Phi(\Delta T), \tag{14}$$

and $$\Gamma = \int_{m\Delta T}^{(m+1)\Delta T} \Phi((m+1)\Delta T-\tau)d\tau g. \tag{15}$$

If Φ in (15) is written as the matrix exponential and expanded in a Maclaurin's series, the integral can readily be evaluated as the infinite series,

$$\Gamma = \sum_{j=0}^{\infty} \frac{(F\Delta T)^j}{(j+1)!} \Delta Tg\ . \tag{16}$$

The expressions for Λ and Γ, (14) and (16), are ideal for direct numerical evaluation when (14) is expanded in a Maclaurin's series.

The Q derived from the continuous design earlier is then inserted in the discrete performance index,

$$J = \sum_{j=0}^{\infty} (x_j^T Q x_j + u_j^2), \tag{17}$$

and the optimal control law for the sampled-data model (13) and the discrete performance index (17) is computed. This optimization problem [K3] is the discrete analog of the continuous one considered throughout this work. This problem results in a constant optimal feedback law computed similarly to the continuous case. If the optimal sampled-data control law is,

$$u_{m+1} = - k^T x_m ,$$

then it is hoped that the closed-loop sampled-data system will perform similarly to the continuous closed-loop system which was originally designed, so that the specifications are complied with by the discrete design as well.

As $\Delta T \to 0$ it is clear that the optimal discrete system will respond identically to the optimal continuous design. For ΔT sufficiently small the discrete design should perform within the specifications. If the sample rate is too slow to faithfully reproduce the desired response, the Q generated from the continuous design will nonetheless provide a basis for design by performance index iteration, as was done for the continuous case in Section 4.3.

By way of illustration, the example of design by performance index iteration will be repeated, this time with the goal of designing a digital compensator which meets the prescribed time domain specifications. The frequency domain specifications (gain and phase margin) will not be investigated since these terms require redefinition for sampled-data systems. Frequency domain analysis for sampled-data systems can generally be accommodated with no difficulty through the use of the "bilinear transforms" [K12].

<u>Example 4.3</u>

The plant of Example 4.2 is to be compensated through the use of a sampled-data controller, with an (at present) unspecified sample rate, so that it meets the following time domain performance cri ia. In response to a step input the system should have

1. Overshoot: $Os \leq 5\%$
2. Rise time (within 90%): $t_r \leq 1$ sec
3. Settling time (within 1%): $t_s \leq 2$ sec.

In the earlier example

$$Q = \begin{bmatrix} 3000 & 0 & 0 \\ 0 & 60 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

was found to result in an optimal control law which met these specifications for the system in companion matrix canonical form.

All that remains is to compute the discrete model of the system and optimal discrete control law corresponding to the Q above. This was done, again using program LQL, with sample period of 0.05, 0.1, and 0.2 seconds. All three sample rates resulted in compensated systems which easily met the specifications. The step response of the optimal discrete system with $\Delta T = 0.1$ is representative of the three trials. This response was within approximately 1% (of the step magnitude) of the response of the final design in Example 4.2 for the first 1.5 seconds and within less than 0.1% thereafter.

The three sample rates were chosen to be near the "rule-of-thumb" value of approximately one-tenth the fastest time constant of the optimal continuous system and are therefore reasonable values. A sampling period of one-half second was also used and resulted in a closed-loop system response very similar to that desired but just failing to meet the specifications.

It would, of course, be rather extravagant to employ a digital compensator to simply feed back a linear combination of the states. However, with the availability of an arithmetic unit (and presumably some memory) in the compensator, the implementation of a Kalman filter to estimate the states in the presence of noise is a relatively minor extension.

As mentioned earlier, the original motivation for this entire study was to develop a technique for retrofitting a digital compensator to a system with a satisfactory continuous compensator without altering system performance. It is now clear how this can be accomplished. The techniques of Section 4.2 are applied to the original continuous system to determine the performance index, if any, minimized by the original configuration.[7] If the original system is optimal, then the design of a digital compensator proceeds as in the example. When sign indefinite Q's are admitted, this is not a severe restriction.

---

[7] Compensator poles should be removed from both open- and closed-loop systems before optimality is tested or the performance index constructed.

CHAPTER V

CONCLUSIONS

5.1 Summary of Results

It often seems in the pursuit of a solution to a specific problem that the many incidental and preliminary results one encounters along the way are of as much lasting value as those originally sought. That apparently has been the case in this study.

The original intent of the research presented here was to develop a scheme to use optimal control theory for the automated design of digital compensators to replace existing continuous ones. This leads to an investigation of the inverse optimal control problem, to determine when a solution to the inverse problem exists and how it is obtained. It was found that the criterion for optimality is a frequency domain result and leads naturally to consideration of the meaning of optimality in a classical (time and frequency domain) context. From the study of the common ground of classical and optimal control theory, the possibility arose of using the linear quadratic loss problem as a computational vehicle for design of linear controllers (continuous and discrete) to classical performance specifications.

The contributions of this research can be summarized as follows:

1. A complete theoretical investigation of the inverse optimal linear regulator problem for quadratic performance indices with positive semidefinite

state weighting matrices and scalar input systems is reviewed and preliminary results for the case where the weighting matrix is allowed to be sign indefinite are given.

2. The equivalence of performance indices for scalar input linear quadratic loss problems is resolved and a procedure for generating the entire equivalence class of cost functions equivalent to a given performance index is developed.
3. Practical numerical methods for determination of system optimality and computation of solutions to the scalar input inverse problem are discussed.
4. Some of the effects of specific elements of the performance index on optimal system performance and pole locations are determined.
5. The problem of designing optimal systems to meet classical performance specifications is encountered and some definitive results obtained.
6. A solution to the problem of designing a sampled-data controller to approximate the performance of continuous controller is given.

5.2 Suggestions for Future Research

Many of the results of this work are quite open-ended. For instance, the goal of using optimal control theory for completely automated design of control systems subject to classical performance

specifications is still far off. Some areas where research can be invested with immediate benefit are outlined below.

1. Although some investigations of the inverse problem for multi-input systems have been made, they are generally very cursory and much work in the spirit of Chapters III and IV here remains to be done.
2. Necessary and sufficient algebraic conditions for the non-existence of conjugate points when the state weighting matrix of the quadratic performance index is sign indefinite are eagerly awaited.
3. More extensive relations between performance index specification and optimal system response need to be developed.
4. Techniques similar to those of Chapter IV can probably be profitably applied to the so-called "model following problem" [A1].

APPENDICES

APPENDIX A

ADDITIONAL PERFORMANCE INDICES ACCOMMODATED BY THE THEORY

A.1 Introduction

As mentioned in the text, several performance indices other than the one specifically discussed, i.e.,

$$J = \int_{t_o}^{t_f} (x^T Qx + u^T Ru)dt, \tag{1}$$

may be formulated which lend themselves well to the interpretations and techniques of this work. To demonstrate this, four quadratic performance indices, which may be viewed as variations of (1), are shown to be reducible to (1) by some rather simple transformations.

None of these performance measures actually extend any of the results presented, except perhaps heuristically. Their value is that they allow the designer to initially specify a performance index which may be more appealing than (1) for one reason or another in a particular case and then to proceed to reduce it to the form of (1) for application of the design techniques presented here.

A.2 Quadratic Performance Index with Exponential Weighting

An interesting variation of (1) discussed by Anderson and Moore [A4] is to include an exponential weighting term in the integrand of the functional (1). That is,

$$J = \int_{t_o}^{t_f} e^{2\alpha t}(x^T Qx + u^T Ru)dt, \quad \text{R positive definite} \tag{2}$$

and the completely controllable system to be optimized is

$$\dot{x} = Fx + Gu \ . \tag{3}$$

If the resulting control law and the system are time invariant, then it is reasonable to assume that the states must decay at least as fast as $e^{-\alpha t}$ in order for (2) to remain bounded. This will indeed be shown to be the case for infinite final time; that is, the real parts of all the closed-loop poles must be less than $-\alpha$.

Let

$$\hat{x} = e^{\alpha t}x \quad \text{and } \hat{u} = e^{\alpha t}u,$$

then

$$\dot{\hat{x}} = e^{\alpha t}\dot{x} + \alpha e^{\alpha t}x,$$

reapplication of the definition of $\hat{x}$ and substitution from (3) for $\dot{x}$ result in

$$\dot{\hat{x}} = e^{\alpha t}(Fx + Gu) + \alpha\hat{x}$$

or

$$\dot{\hat{x}} = (F + \alpha I)\hat{x} + G\hat{u} \ . \tag{4}$$

Then (2) can be rewritten as

$$J = \int_{t_o}^{t_f} (\hat{x}^T Q\hat{x} + \hat{u}^T R\hat{u})dt$$

which along with (4) is an optimal regulator problem in the usual sense and has the optimal control law (assuming no conjugate points),

$$\hat{u} = - \hat{K}^T(t)\hat{x}.$$

Returning the control law to the original coordinate system,

$$u = - e^{-\alpha t} \hat{K}^T(t) e^{\alpha t} x$$

$$= - \hat{K}^T(t) x \ ,$$

and if t_f is allowed to become infinite, $\hat{K}$ becomes a constant control. If the optimal system is stable, $\hat{x}$ approaches zero as t becomes infinite; hence, x decays at least as fast as $e^{-\alpha t}$ and the original interpretation of (2) is shown to be correct.

In summary, the performance index with exponential weighting can be accommodated by a redefinition of the plant which has the effect of shifting the plant poles by $-\alpha$. If α is positive, this can be thought of as giving the optimization problem, which tends to generate closed-loop poles to the left of the plant poles, a "head start" of α.

A.3 Quadratic Performance Index with Cross-Product

Consider now the generalization of (1) which includes cross-products of the states and inputs, i.e.,

$$J = \int_{t_o}^{t_f} (x^T Q x + 2x^T A u + u^T R u) dt, \tag{5}$$

where A is an $n \times m$ real matrix, m is the number of inputs, and R is positive definite and symmetric.

As before, a redefinition of the model will reduce this variation to the form of (1), in this case only the control need be modified. Let

$$\hat{u} = u + B^T x,$$

where B is an $n \times m$ real matrix, then the integrand of J is

$$x^TQx + 2x^TA(\hat{u}-B^Tx) + (\hat{u}-B^Tx)^TR(\hat{u}-B^Tx)$$

$$= x^T(Q-2AB^T+BRB^T)x + 2x^T(A+BR)\hat{u} + \hat{u}^TR\hat{u}.$$

The problem will reduce to the usual form if

$$A = BR;$$

since R is positive definite it is also non-singular and the required B is

$$B = AR^{-1} \quad . \tag{6}$$

The optimization problem with performance index (5) may be solved by replacing the Q of the usual performance index (1) with

$$\hat{Q} = Q - 2AB^T + BRB^T$$

from (5); substitution of (6) for B simplifies this considerably to

$$\hat{Q} = Q - AR^{-1}A^T \quad .$$

Then, if the customary optimization problem with $\hat{Q}$ has no conjugate points, it will have the optimal control law, $\hat{u} = -\hat{K}^Tx$, and the optimal control for (5) is

$$u = -(\hat{K} + AR^{-1})^Tx. \tag{7}$$

Clearly, there is no increase in generality if A of (5) is chosen to be non-null, since no additional optimal systems would be admitted. Formulation (5) permits a change in the appearance of the optimization problem with no effect on the theory.

A.4 Quadratic Performance Indices with Derivative Weighting

Two variations of derivative weighting performance indices will now be considered. First, a quadratic form composed of the (first) derivatives of the states will be added to (1); that is,

$$J = \int_{t_o}^{t_f} (x^T Qx + \dot{x}^T C\dot{x} + u^T Ru)\,dt, \tag{8}$$

where C is a symmetric $n \times n$ real matrix. Upon substitution from (3) for $\dot{x}$ the integrand becomes

$$x^T Qx + (Fx+Gu)^T C\,(Fx+Gu) + u^T Ru$$

and simplifies to

$$x^T(Q+F^T CF)x + 2x^T F^T C\,Gu + u^T(R+G^T CG)u.$$

This integrand is of the same form as the performance index of the last section and hence can possibly be put in the customary form by a simple redefinition of u. In this case the required B (6) is

$$B = F^T CG(R+G^T CG)^{-1} \tag{9}$$

and may not exist. If the necessary transformation exists

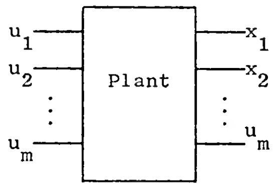
$$R + G^T CG$$

must be non-singular. This must also be non-singular for the equivalent R of the resulting performance index to be positive definite; thus, if the optimization problem with (8) has a solution, the transformation (9) must exist. Then (8) is again different from (1) in appearance only.

The second variation of derivative weighting is to include a quadratic form of the derivative of the control; i.e.,

$$J = \int_{t_o}^{t_f} (x^T QX + \dot{u}^T D\dot{u} + u^T Ru)dt, \tag{10}$$

where D is a real $m \times m$ symmetric matrix. This case is handled by renaming the control and adding m (one per input) additional states. Let $\hat{u} = \dot{u}$, then graphically the original configuration



is changed so that the system to be optimized is

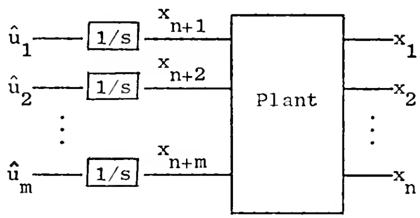


Figure A.1 Modified Plant

Define the state model for the modified (n + m order) plant as

$$\dot{\hat{x}} = \hat{F}\hat{x} + \hat{G}\hat{u},$$

where $$\hat{F} = \begin{bmatrix} F & G \\ 0 & 0 \end{bmatrix}, \qquad \hat{G} = \begin{bmatrix} 0 \\ I_m \end{bmatrix} \tag{11}$$

and $\hat{x}$ has components as in Figure A.1. Then, if $\hat{Q}$ and $\hat{R}$ in

$$J = \int_{t_o}^{t_f} (\hat{x}^T\hat{Q}\hat{x} + \hat{u}^T\hat{R}\hat{u})dt$$

are chosen as (12)

$$\hat{Q} = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \qquad \text{and} \qquad \hat{R} = D,$$

the solution to the optimization problem with derivative weighting (10) may be obtained from the customary problem. Assume that the optimization problem with (11) and (12) has a solution (no conjugate points) which is

$$\hat{u} = - \hat{K}\ \hat{x}\ . \tag{13}$$

By partitioning $\hat{K}$ into two matrices,

$$\hat{K} = \begin{array}{c} \left[\begin{array}{c} K_1 \\ \hdashline K_2 \end{array}\right] \end{array} \begin{array}{l} n \\ m \end{array},$$

and using the definition for $\hat{x}$ from (11), the control law (13) can be rewritten as

$$\hat{u} = - \int_{t_o}^{t} (K_1^T x + K_2^T\hat{u})dt.$$

In this form a realization for the optimal control law in terms of the original plant is clearly:

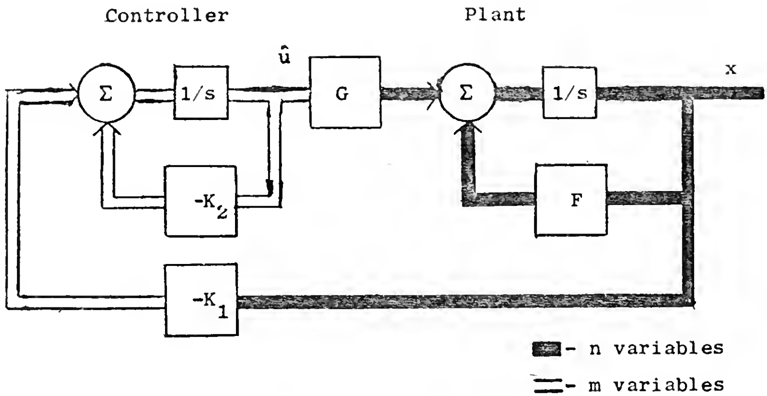


Figure A.2 Synthesis of Optimal Control for Performance Index with Control Derivative Weighting

This case is somewhat different from the other performance indices which have been considered in that the optimal control must be synthesized with a dynamic compensator. However, it is obvious from the modified form of the plant and the performance index that it also is but a variation of the original problem.

APPENDIX B

PROOF OF SUFFICIENCY OF THEOREM 3.1

To prove the sufficient part of Theorem 3.1 it must be shown that for a scalar input completely controllable system,

$$\dot{x} = Fx + gu, \tag{1}$$

and stable completely observable control law,

$$u = - k^T x; \tag{2}$$

if

$$\left|1+k^T\Phi(j\omega)g\right|^2 \geq 1 \qquad \Phi(s) = (sI-F)^{-1}, \tag{3}$$

for all real ω, then the control law (2) is optimal.

Since, by the hypothesis, the system (1) is completely controllable it can be taken without loss of generality to be in companion matrix canonical form (Section 3.4). Then by Lemma 3.1,

$$S(s) = \varphi(s)\Phi(s)g = \begin{bmatrix} 1 \\ s \\ s^2 \\ \vdots \\ s^{n-1} \end{bmatrix}, \text{ where } \varphi(s) = \det(sI-F) \tag{4}$$

and $$k^T\Phi(s)g = \frac{k_1 + k_2 s + \dots + k_n s^{n-1}}{\varphi(s)} = \frac{K(s)}{\varphi(s)} .$$

Application of this expression to (3) results in,

$$\left|1+k^T\Phi(j\omega)g\right|^2 = \left|\frac{\varphi(j\omega) + K(j\omega)}{\varphi(j\omega)}\right|^2 \geq 1; \tag{5}$$

then from the assumption of companion matrix canonical form for (1), it follows that

$$\varphi(j\omega) + K(j\omega)$$

is the closed-loop characteristic polynomial, which will be denoted $\varphi_k(j\omega)$.

Inequality (5) can now be rewritten as

$$\Psi(\omega) = |\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2 \geq 0,$$

displaying the non-negativity of Ψ which allows it to be factored (Lemma 3.2) as

$$\Psi(\omega) = \psi(-j\omega)\psi(j\omega),$$

where $\psi(s)$ is a real Hurwitz polynomial and can be represented as the vector product

$$\psi(s) = p^T S(s), \quad p^T = [p_1, p_2, \ldots, p_n] .$$

From the proof of the necessary part of Theorem 3.1, if k is an optimal control law, then

$$|1+k^T\Phi(j\omega)g|^2 = 1 + g^T\Phi^T(-j\omega)Q\Phi(j\omega)g,$$

which can be rewritten in the same manner as (5) to result in

$$\Psi(\omega) = |\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2 = g^T\Phi^T(-j\omega)Q\Phi(j\omega)g. \tag{6}$$

Again evoking relation (4), it is clear that

$$Q = pp^T \tag{7}$$

will satisfy (6).

If it can now be shown that the use of this weighting matrix in an optimization with (1) will result in k, the theorem is proved.

Suppose that the performance index,

$$J = \int_0^\infty (x^T pp^T x + u^2)dt,$$

subject to the system (1) is minimized by the control law, $k_1 \neq k$, and denote the resulting closed-loop characteristic polynomial as $\varphi_{k_1}(s)$. By (6) and the definition of Q in (7)

$$|\varphi_{k_1}(j\omega)|^2 = |\varphi_k(j\omega)|^2; \tag{8}$$

then if k_1 is shown to be a stable control law, the unique Hurwitz spectral factor (Lemma 3.2) of both sides of (8) will be identical and $k_1 = k$.

The optimal control law k_1 will be stable if the pair [F,p] is completely observable (Theorem 2.5). Suppose they are not completely observable. Then the numerator and denominator of

$$p^T \Phi(s) g = \frac{\psi(s)}{\varphi(s)}$$

will have a common factor [K5] which must have zeros with negative real parts since ψ is Hurwitz. By rewriting (6) as

$$|\varphi_k(j\omega)|^2 = |\psi(j\omega)|^2 + |\varphi(j\omega)|^2,$$

it is clear that $\varphi_k(s)$ must contain the same common factor, since k is a stable control law and $\varphi_k(s)$ must have only zeros with negative real parts. Then from the observation in (5) that

$$\varphi_k(s) = \varphi(s) + K(s),$$

K must also contain the same common factor as φ and φ_k. But this contradicts the stipulation in the hypothesis that k be completely observable. Hence the pair [F,p] are completely observable and the theorem is proven.

APPENDIX C

THE INVERSE PROBLEM - NUMERICAL DETAILS

C.1 Introduction

In this appendix the numerical methods developed during the course of this research for computation of solutions to the inverse problem are discussed. Specifically, the algorithms presented here deal with the problems of:

1. computing $\Psi(\omega)$ for a given system configuration,
2. determining the non-negativity of Ψ, and
3. the computation of the Hurwitz spectral factor of Ψ once it has been determined to be non-negative.

These are essentially the procedures required to determine if a completely controllable system and stable scalar control law are optimal with respect to a quadratic performance index with Q positive semidefinite and when it has been ascertained to be optimal, computing a unity rank Q which forms a performance index minimized by the optimal system. The additional numerical operations required to investigate the optimality of a closed-loop system with respect to a performance index which has a Q not positive semidefinite are very straightforward (Section 4.2) and nothing would be gained by including them here.

C.2 Computation of Magnitude - Square Polynomials

The first step in the determination of the optimality of a given system configuration is computation of the real even polynomial,

$$\Psi(\omega) = |\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2 .$$

This necessitates the calculation of the coefficients of the magnitude-square polynomials for the open- and closed-loop characteristic polynomials; the algorithm presented in this section is an efficient technique for the computation of these coefficients. The method presented is unique in that it does not require complex arithmetic and computes only the non-zero terms.

Identification MAGSQ

SUBROUTINE MAGSQ (X, Z, NDIM)

Purpose

To compute the magnitude-square polynomial

$$Z(\omega^2) = |X(j\omega)|^2 = X(j\omega)\ X(-j\omega)$$

of a given real polynomial.

Method

Let $Z(\omega^2) = X(j\omega)\ X(-j\omega)$

and $$Z(\omega^2) = \sum_{i=0}^{N} z_i \omega^{2i} \qquad X(s) = \sum_{i=0}^{N} x_i s^i ,$$

then since $Z(\omega^2)$ is even only the even power coefficients need be computed, that is, $z_{i+k} = \sum_{i=0}^{N} \sum_{k=0}^{N} (j)^i\ x_i (-j)^k\ x_k$

such that $i + k$ is even. Each coefficient of Z is then,

$$z_\ell = \sum_{\substack{i+k=\ell \\ \text{even}}} (-1)^k\ (j)^{i+k}\ x_i x_k$$

and because $i + j$ is an even integer,

$$(j)^{i+k} = (-1)^{(i+k)/2}$$

and the previous expression may be rewritten

$$z_\ell = \sum_{\substack{i+k=\ell \\ \text{even}}} (-1)^k \, (-1)^{(i+k)/2} \, x_i x_k \, .$$

Then the expression used for computing the coefficients of Z is

$$z_\ell = \sum_{\substack{i+k=\ell \\ \text{even}}} (-1)^{(i+3k)/2} \, x_i x_k ,$$

which requires no complex arithmetic and only the $N+1$ non-zero terms are computed.

Usage

CALL MAGSQ (X, Z, NDIM)

Input

X - vector of polynomial coefficients stored from smallest to largest power terms (single or double precision, see Remarks)

NDIM - number of coefficients of X and Z, i.e., $N+1$

Output

Z - vector of magnitude-square polynomial coefficients stored from smallest to largest even power terms, i.e., nth term of vector is coefficient of $\omega^{2(n-1)}$ (single or double precision, see Remarks)

Remarks

Coefficient arrays X and Z are presently double precision but may be made single precision by placing a "C" in column one of card 12 as indicated below:

```
C    DOUBLE PRECISION X,Z
```

C.3 Test for Non-negativity of Real, Even Polynomials

After the computation of the magnitude-square of the open- and closed-loop characteristic polynomials, optimality of the system with respect to a quadratic performance index with positive semidefinite state weighting matrix can be determined from the sign definiteness of $\Psi(\omega)$. This test is provided by subprogram NONNEG and supporting subroutine NSIGNV; both subroutines are Fortran integer functions. The heart of the procedure is a novel method for computation of a normalized Routh Array and the only known implementation of Šiljak's test for non-negativity.

Identification

NONNEG

FUNCTION NONNEG (P,NDIM,IPRT)

Purpose

To verify non-negativity of even real polynomials, that is, to determine the optimality with respect to a quadratic performance index (with positive semidefinite state weighting matrix) of a constant coefficient linear dynamical system with a given scalar feedback control law.

Method

The requirement that the real even polynomial, $\Psi(\omega)$, be non-negative for all real ω is equivalent (Section 4.2) to requiring that $\Psi(\omega^2)$ possess no positive real zeros of odd multiplicity. Hence this algorithm relies on properties related to root location to determine non-negativity.

The procedure begins with four simple necessary tests (one sufficiency test is made incidental to a necessary test); if all are

successful (and the sufficiency test fails) a necessary and sufficient condition for positive definiteness is tested. If this test fails, a final necessary and sufficient test for non-negativity is made. In general, the philosophy of this procedure is to attempt to resolve the sign definiteness of the given polynomial with the simplest test possible and to resort to sophisticated techniques only if all else fails. The following necessary conditions are first tested (in order):

$$\Psi(\omega^2) = \sum_{i=1}^{N+1} a_i \omega^{2(i-1)}$$

1. $a_1 \geq 0$, i.e., $\Psi(0) \geq 0$
2. $a_{N+1} \geq 0$, i.e., $\lim_{\omega \to \infty} \Psi(\omega) \geq 0$
3. $\Psi(\omega^2)$ must have an even number of sign variations between successive terms (a_i's) in order to have an even number of positive real zeros by Decartes' rule of signs. (If there are zero sign variations the coefficients are all non-negative and this condition is also sufficient.) The test for an even number of sign variations is actually implicit in 1 and 2 (above); however, it is made as a convenient test for zero sign variations and as a check for possible numerical problems.

If all the preceding tests on necessary conditions are passed, a normalized Routh Table [G4] is constructed by normalizing the rows to have a maximum modulus of unity and the following sufficiency test [J2] is made:

A real, even polynomial $\Psi(\omega^2)$ has no positive real zeros (hence is positive for all real ω) if and only if the number of sign variations

between successive coefficients is even (3 above) and the number of sign variations in the first column of the Routh Table is at least N-1 (where 2(N-1) is the highest power in $\Psi(\omega)$).

If all of the necessity tests have been passed but no sufficiency condition has been satisfied, a modification of Šiljak's necessary and sufficient test for non-negativity is made as follows [S3, K10]:

1. Number the rows of the Routh Array starting at the top $r = 1,2,\ldots,2N+1$.
2. Denote each row preceding a zero row by $r_\nu, \nu=1,2,\ldots,m$ where the first row is r_o and the last row r_{m+1}.
3. V_ν is the number of sign variations between two consecutive zero rows r_ν and $r_{\nu-1}$.
4. $\Pi_\nu = 2r_\nu - r_{\nu-1} - r_{\nu+1} - V_\nu + V_{\nu+1} \qquad \nu = 1,2,\ldots,m$

   $\Pi_{m+1} = r_{m+1} - r_m - V_{m+1}$.
5. If the sum of Π_ν over all odd ν is zero, then $\Psi(\omega^2)$ is non-negative.

Usage

ITEST = NONNEG (PSI,NDIM,IPRT)

Input

PSI - Vector of coefficients of real, even polynomial to be tested (only even power terms are included).

NDIM - Dimension of polynomial, i.e., N+1.

IPRT - Flag to control printing of Routh Table (if computed) and results of non-negativity test.

IPRT = 0 results in no printing.

Output

NONNEG is set to one of the following integer values depending on the results of the non-negativity test.

0 - Error in NONNEG, no conclusion about non-negativity of PSI.

1 - Constant term of PSI is negative, PSI is not non-negative.

2 - Highest power term of PSI which is non-zero is negative, PSI is not non-negative.

3 - PSI has an odd number of sign variations between consecutive coefficients (Descartes' rule of signs), PSI is not non-negative.

4 - PSI passes all of the necessity tests but fails Šiljak's test, PSI is not non-negative.

5 - PSI has all non-negative coefficients, PSI is positive definite.

6 - PSI has an even number of sign variations and its Routh Table has at least N-1 sign variations in the first column, PSI is positive definite.

7 - PSI passes Šiljak's test, PSI is non-negative.

Restrictions

Calls function subprogram NSIGNV.

Identification NSIGNV

FUNCTION NSIGNV (X,B,L)

Purpose

To determine the number of sign variations between consecutive terms of a real vector.

Method

The L elements of the input vector X are examined for sign, a corresponding ± 1 is placed into working vector B and zero elements of X are disregarded. The L-M (where M is the number of zero elements of X) elements of B are sequentially multiplied and every $B(I-1) \cdot B(I) < 0$ is recorded as a sign variation.

Usage

NSIGN = NSIGNV (X,B,L)

Input

X - real vector to be tested, at least L elements long

B - real working vector, at least L elements long

L - number of terms of X to be tested

Output

NSIGNV is set to the total (integer) number of sign variations between L consecutive terms of the vector X.

C.4 Spectral Factorization

If $\Psi(\omega)$ is non-negative, the system configuration being tested is optimal and by Theorem 3.2 a performance index with a unity rank positive semidefinite state weighting matrix can always be constructed. This weighting matrix is formed from the outer product of the vector of coefficients of the Hurwitz spectral factor of $\Psi(\omega)$ ordered with constant term first (Section 3.5). This section presents an efficient and accurate method for the extraction of Hurwitz spectral factors of non-negative real even polynomials. The procedure is somewhat unusual

in that no roots of the polynomials are explicitly computed. This procedure is implemented by subprogram SPFCTR and supporting subroutine QFACT.

Identification SPFCTR

SUBROUTINE SPFCTR (PSI,FACTR,NDIM,IER)

Purpose

To compute the Hurwitz spectral factor of a real, even, non-negative polynomial.

Method

A factor $\psi(s)$ of the input polynomial $\Psi(\omega)$ is to be computed so that

$$\Psi(\omega) = \psi(j\omega)\ \psi(-j\omega) = \sum_{i=0}^{N+1} a_i \omega^{2i}$$

and $\psi(s)$ has all left half-plane zeros.

First, the formal substitution of $\sigma = \omega^2$ into $\Psi(\omega^2)$ reduces the order of Ψ from 2N to N and maps the zeros of $\psi(s)$ (and consequently zeros of $\Psi(\omega)$) from

$$\pm\alpha \pm j\beta, \text{ to: } (\alpha^2 - \beta^2) \pm j2\alpha\beta\ .$$

An approximation $P(\sigma)$, to a quadratic factor of $\Psi(\sigma)$ is then generated using Bairstow's method [H3], Ψ is divided by P and its order reduced by two. P becomes a factor F of the spectral factor after its coefficients have been altered to reflect the inverse of the mapping described above. This is done by computing the discriminant of P and by examining its sign along with the signs of the coefficients of P to determine the necessary transformation to compute the coefficients of F.

The steps in the preceding paragraph are repeated and the quadratic factors F of the spectral factor are multiplied together to produce the complete spectral factor. This method avoids the inaccuracies involved in determining the 2N roots of $\Psi(\omega)$ and constructing $\psi(s)$ from the left half-plane zeros.

Usage

CALL SPFCTR (PSI,FACTR,NDIM,IER)

Input

PSI - Vector of coefficients of the polynomial which is to be factored, only even power terms are stored, i.e., nth term of the vector is the coefficient of $\omega^{2(n-1)}$ (double precision).

NDIM - Number of coefficients of PSI, i.e., N+1.

Output

FACTR - Vector of coefficients of the Hurwitz spectral factor of PSI (if non-negative) stored from smallest to largest power terms (double precision).

IER - Integer error code:

0 - no error

1 - maximum number of attempts at finding quadratic factors of $\Psi(\sigma)$ exceeded (presently, 2(N+1) - factorization aborted.

2 - order of input polynomial has been reduced by roundoff or overflow - factorization aborted.

3 - remaining linear factor of $\Psi(\sigma)$ has a positive zero, no real spectral factor can be generated - factorization aborted.

Restrictions

If the input polynomial is not non-negative, a real spectral factor does not exist; nonetheless, the algorithm may still produce an attempt at a spectral factor, without detecting an error. The results of this procedure can only be relied upon if the non-negativity of the input polynomial is insured.

Calls subprogram QFACT.

Identification QFACT

SUBROUTINE QFACT (C,IC,Q,LIM,DNORM,IER)

Purpose

To compute and divide out a quadratic factor of a real polynomial using Bairstow's method.

Method

Let the input polynomial to the algorithm be:

$$C(s) = \sum_{i=0}^{N} c_i s^i$$

and an initial guess for a quadratic factor be: $s^2 - vs - w$. By division, a quotient and a remainder are computed,

$$C(s) = (s^2 - vs - w)\, B(s) + Rs + T,$$

if $R = T = 0$, the initial guess was correct and it is a quadratic factor. If $R \neq T \neq 0$, then a sequence of v's and w's is generated so that in the limit $R = T = 0$. Bairstow's method is simply the simultaneous solution of

$$R(v, w) = 0$$

and

$$T(v, w) = 0$$

by Newton's method [H3].

Once a satisfactory approximation to a quadratic factor has been obtained, C is replaced by the quotient and the order is reduced by two.

Usage

CALL QFACT (C,IC,P,LIM,DNORM,IER)

Input

C - vector of coefficients of the polynomial to be factored, stored from smallest to largest power term, i.e., nth term of vector is coefficient of s^{n-1} (double precision)

IC - number of coefficients of C.

P - vector of coefficients of an initial guess for a quadratic factor (on input): $s^2 + P(2)s + P(1)$ (double precision).

LIM - integer specifying the maximum number of iterations allowed.

Output

C - vector of coefficients of the polynomial with quadratic factor removed, the last coefficient is now one (double precision).

IC - number of coefficients of reduced C, i.e., value on input less two.

P - approximation of quadratic factor: $s^2 + P(2)s + P(1)$ (double precision).

DNORM - value by which coefficients of C on input have been normalized, if other than one, DNORM = C(IC) (double precision).

IER - error code set to one of the following integer values:

0 - no error encountered

1 - no convergence within LIM iterations

-1 - input polynomial is constant or undefined - or an overflow occurred during normalization

-2 - input polynomial is of degree one

-3 - no further refinement of the approximation to a quadratic factor possible; division by zero, overflow, or an initial guess outside of the region of convergence.

Remarks

Some of the FORTRAN code was taken from parts of subroutine DPQFB of the IBM System/360 Scientific Subroutine Package [I1, pp. 193-197].

C.5 Sample Program

OPTI is a sample main program which demonstrates how a given design configuration may be tested for optimality with respect to a quadratic performance index and how appropriate performance indices are computed if they exist.

The program accepts as inputs the coefficients of the closed-loop and open-loop characteristic polynomials of the system under investigation. The stability of the closed-loop characteristic polynomial is not tested. It is assumed that the closed-loop characteristic polynomial resulted from a design technique which insured stability; if the closed-loop system is not stable, the results of program OPTI are not supportable.

The even polynomial,

$$\Psi(\omega) = |\varphi_k(j\omega)|^2 - |\varphi(j\omega)|^2,$$

is formed and tested by subroutine NONNEG for system optimality.

If the system is not optimal for a positive semidefinite state weighting matrix, that is, $\Psi(\omega) \not\geq 0$, the program terminates analysis of the present system and interrogates the input file for further instructions.

If the system is determined to be optimal, corresponding positive semidefinite state weighting matrices (Q) of the performance index,

$$J = \int_0^\infty (x^T Q x + u^2)\, dt,$$

are computed. If $\Psi(\omega^2)$ has all non-negative coefficients, the diagonal Q is printed. In either case, the unity rank factor of Q, that is, h where $Q = hh^T$, is computed by spectral factorization using subroutine SPFCTR and printed.

The accuracy of the factorization is checked by performing the product,

$$\pi(\omega) = \psi(j\omega)\psi(-j\omega),$$

(where ψ is the approximate spectral factor of Ψ) and determining the average magnitude difference in the coefficients of π and Ψ. The results of the accuracy test are also printed.

Several tests for numerical difficulties are made in the course of the analysis and, if a problem arises, an appropriate explanation is printed and the input is queried for instructions.

<u>Input</u>

Each input card image must be in the following format:

Field	Columns	Type	Use
1	1-3	Alpha	Instruction flag
2	4,5	Integer	Coefficient number
3	6-30	Real*	Coefficient

*Stored internally in double precision and read on format D25.16.

Field 1 contains (left justified) one of the following instruction flags:

P - The data on this card (and all following cards until another flag is encountered) describe a coefficient of polynomial φ (open-loop characteristic polynomial).

PK - Same as P except identifies coefficients of φ_k (closed-loop characteristic polynomial).

*GO - Signals the end of data for a given system, commands the program to begin analysis.

END - All program execution is to terminate.

Field 2 contains an integer (right justified), i, indicating that the datum in field 3 is the coefficient of the (i-1) power term of the polynomial indicated in field 1 of this or a previous card.

Field 3 contains a coefficient value in (either single or double precision) FORTRAN real number format. Only non-zero coefficients need be entered (i.e., the coefficients are zeroed prior to reading each case).

Sample Case

The output of a sample case is included in Section C.6.

C.6 Subprogram Listing

In order to facilitate reference to this program, the listing has been divided into functional groups of subprograms. These functional groups are indexed below.

I. Test for Non-negativity of Real, Even Polynomials

1. NONNEG
2. NSIGNV

II. Spectral Factorization

1. SPFCTR
2. QFACT

III. Computation of Magnitude - Square Polynomials

1. MAGSQ

IV. Sample Program

1. OPTI
2. Sample Problem

```
      FUNCTION NONNEG(A2K,NDIMA,IPRT)
C     FUNCTION NONNEG DETERMINES THE NON-NEGATIVITY OF EVEN REAL POLYS.
C
C        NONNEG RETURNS THE FOLLOWING CODESO
C        0 - ERROR IN NONNEG
C        1 - CONSTANT TERM NEGATIVE
C        2 - HIGHEST POWER (NONZERO) TERM NEGATIVE
C        3 - POLYNOMIAL HAS ODD NUMBER OF SIGN VARIATIONS
C        4 - POLYNOMIAL HAS EVEN NO. OF SIGN VARS. HAS LESS THAN N-1
C            SIGN VARS. IN 1ST. COL. OF ROUTH TABLE FAILS SILJAK'S TEST
C        5 - POLYNOMIAL COEFICIENTS ALL POSITIVE
C        6 - POLYNOMIAL HAS EVEN NO. OF SIGN VARS. ROUTH TABLE HAS AT
C            LEAST N-1 SIGN VARIATIONS IN 1ST. COLUMN
C        7 - POLYNOMIAL HAS EVEN NO. OF SIGN VARS. ROUTH TABLE DOES NOT
C            HAVE N-1 SIGN VARS. IN 1ST. COLUMN PASSES SILJAK'S TEST
C                                                    2(N-1)
C     A2K - INPUT POLYNOMIAL COEFS.  A2K(N) = COEF. OF W
C     NDIMA - NUMBER OF NONZERO COEFICIENTS
C     IPRT - FLAG TO INDICATE WHETHER THE NORMALIZED ROUTH TABLE AND
C             THE RESULTS OF NONNEG ARE TO BE PRINTED (IPRT.NE.0)
C                                               J.M. ELDER  4/20/71
      DOUBLE PRECISION DR1,DR2,DTEMP,R1M,R2M,SIGNR,RTSQ
      DIMENSION A2K(1),R(40,20),R1(1),B(1),JROW(40),NV(1),
     1DR1(20),DR2(20),DTEMP(20),IOUT(12,8)
      COMMON R,DR1,DR2,DTEMP,R1M,R2M
      EQUIVALENCE (R(1,1),R1(1)),(R(1,2),NV(1)),(R(1,3),B(1))
      DATA IOUT/4H.NOT,4HNO C,4HONCL,4HUSIO,4HN PO,4HSSIB,4HLE, ,4HERRO,
     14HR IN,4H NON,4HNEG ,4H    ,4H.NOT,4HCONS,4HTANT,4H TER,4HM OF,
     24H POL,4HYNOM,4HIAL ,4HNEGA,4HTIVE,2*4H    ,4H.NOT,4HHIGH,4HEST ,
     34HPOWE,4HR CO,4HEFFI,4HCIEN,4HT NE,4HGATI,4HVE  ,2*4H    ,4H.NOT,
     44HPOLY,4HNOMI,4HAL H,4HAS O,4HDD N,4HO. O,4HF SI,4HGN V,4HARIA,
     54HTION,4HS   ,4H.NOT,4HPASS,4HES N,4HECES,4HSITY,4H, BU,4HT FA,
     64HILS ,4HSOFF,4HICIE,4HNCY ,4HTEST,4H....,4HPOLY,4HNOMI,4HAL H,
     74HAS A,4HLL P,4HOSIT,4HIVE ,4HCOEF,4HFICI,4HENTS,4H    ,4H....,
     84HROUT,4HH TA,4HBLE ,4HHAS ,4H N S,4HIGN ,4HVARS,4H. IN,4HFIRS,
     94HT CO,4HL.  ,4H....,4HPOLY,4HNOMI,4HAL P,4HASSE,4HS SU,4HFFIC,
     A4HIENC,4HY TE,4HST  ,2*4H    /
      DATA IMIN/1H-/,IZER/1H0/,IPLUS/1H+/
   10 FORMAT(///,20X,22HNORMALIZED ROUTH ARRAY/6X,6H ROW  ,5X,
     115HNORMALIZED ROWS/1H ,70(1H-))
   20 FORMAT(6X,1H ,I3,2H  ,1A1,1P8E11.2)
   30 FORMAT(6X,1H ,I3,2H  ,1A1,1P8E11.2,(/6X,1H ,4X,2H  ,1P8E11.2))
   40 FORMAT(3H *R,I2,2H  ,I3,2H  ,1A1,1P8E11.2)
   50 FORMAT(3H *R,I2,2H  ,I3,2H  ,1A1,1P8E11.2,
     1(/6X,1H ,4X,2H  ,1P8E11.2))
   60 FORMAT(28H * ROWS PRECEEDING ZERO ROWS/66H  (ZERO ROWS HAVE BEEN R
     1ESOLVED BY DIFFERENTIATING PRECEEDING ROW))
   70 FORMAT( 49X,1H2/32H  POLYNOMIAL HAS BEEN MULTIPLIED,
     119H BY THE FACTOR  S +,F7.4)
   80 FORMAT(/10X,1H2/5X,12HPSI(W ) IS..,1A4,18H POSITIVE SEMI DEF,
     1 7HINITE, ,11A4)
      ICHNG=0
C     A2K(1)...CONSTANT TERM MUST BE NON-NEGATIVE FOR NON-NEGATIVITY
      IF(A2K(1))90,100,100
   90 NONNEG=1
      GO TO  630
C     REDUCE DIMENSION IF HIGHEST ORDER TERMS NEGLIGIBLE
  100 IF(ABS(A2K(NDIMA))-1.0E-50)110,110,120
  110 NDIMA=NDIMA-1
      GO TO  100
C     TEST FOR HIGHEST ORDER TERM POSITIVE
  120 NDIM=NDIMA
      IF(A2K(NDIM))130,620,140
  130 NONNEG=2
```

```
      GO TO  630
C     CHECK FOR POSITIVITY USING DESCARTES' RULE OF SIGNS
C        DETERMINE NUMBER OF SIGN VARIATIONS
  140 ICHNG=NSIGNV(A2K,B,NDIM)
      IF(ICHNG)620,150,160
  150 NONNEG=5
      GO TO  630
  160 FICHNG=FLOAT(ICHNG)/2.0+1.0E-4
C     TEST FOR ODD NUMBER OF SIGN VARIATIONS
      IF(2*INT(FICHNG)-ICHNG)170,180,620
  170 NONNEG=3
      GO TO  630
C     COMPUTE ROUTH ARRAY
  180 DO  190 I=1,NDIM
  190 DTEMP(I)=DBLE(A2K(I))
C     FILL FIRST TWO ROWS
      NCOL10=0
  200 IZERO=1
      NROW(1)=1
      SIGNR=1.0D0
      J=NDIM+1
      R1M=0.0D0
      R2M=0.0D0
      DO  240 I=1,NDIM
      J=J-1
      DR1(J)=SIGNR*DTEMP(I)
      DR2(J)=2.0D0*DBLE(FLOAT(I-1))*DR1(J)
      IF(DABS(DR1(J))-R1M)220,220,210
  210 R1M=DABS(DR1(J))
  220 IF(DABS(DR2(J))-R2M)240,240,230
  230 R2M=DABS(DR2(J))
  240 SIGNR=-SIGNR
C     NORMALIZE THE FIRST TWO ROWS
      DO  250 J=1,NDIM
      DR1(J)=DR1(J)/R1M
      DR2(J)=DR2(J)/R2M
      R(1,J)=SNGL(DR1(J))
  250 R(2,J)=SNGL(DR2(J))
C     COMPUTE REST OF ARRAY
      NEND1=2*NDIM-1
      DO  400 I=3,NEND1
      NEND=NDIM-INT((FLOAT(I)-1.0)/2.0+1.0E-4)
      SIGNR=1.0D0
      IF(DR2(1))260,260,270
  260 SIGNR=-1.0D0
  270 R1M=0.0D0
      DO  290 J=1,NEND
      DTEMP(J)=SIGNR*(DR2(1)*DR1(J+1)-DR1(1)*DR2(J+1))
      IF(DABS(DTEMP(J))-R1M)290,290,280
  280 R1M=DABS(DTEMP(J))
  290 CONTINUE
C     CHECK FOR ZERO ROWS IN ARRAY
      IF(R1M -1.0D-12)300,300,340
C     NOTE THAT ZERO ROW HAS BEEN ENCOUNTERED AND STORE ITS ROW NO. -1
  300 IZERO=IZERO+1
      NROW(IZERO)=I-1
C     RESOLVE ZERO ROWS
      NPWR=2*NDIM-I
      R1M=0.0D0
      DO  330 J=1,NDIM
      IF(NPWR)340,340,310
  310 DTEMP(J)=DBLE(FLOAT(NPWR))*DR2(J)
      IF(DABS(DTEMP(J))-R1M)330,330,320
  320 R1M=DABS(DTEMP(J))
  330 NPWR=NPWR-2
C     NORMALIZE ROW
```

```
  340 DO  350 J=1,NEND
      DR1(J)=DR2(J)
      DR2(J)=DTEMP(J)/RIM
  350 R(I,J)=SNGL(DR2(J))
C     TEST FOR ZERO IN FIRST COLUMN
      IF(DABS(DR2(1))-1.0D-12)360,360,390
C     MULTIPLY IN AN OFFSETTING FACTOR AND BEGIN TABLE OVER
  360 IF(NCOL10-10)370,620,620
  370 RTSQ=1.111111111111111D0+DBLE(FLOAT(NCOL10))
      NDIM=NDIMA+1
      DTEMP(1)=DBLE(A2K(1))*RTSQ
      DTEMP(NDIM)=DBLE(A2K(NDIMA))
      DO  380 I=2,NDIMA
  380 DTEMP(I)=RTSQ*DBLE(A2K(I))+DBLE(A2K(I-1))
      NCOL10=NCOL10+1
      GO TO  200
  390 NCOL10=0
      DR1(J+1)=DR2(J+1)
      DR2(J+1)=0.0D0
  400 R(I,J+1)=0.0
C     INCLUDE LAST ROW
      IZERO=IZERO+1
      NROW(IZERO)=NEND1
      IF(IPRT)410,530,410
C     PRINT NORMALIZED ROUTH TABLE
  410 WRITE(6,  10)
      KLESS1=0
      K=1
      DO  520 I=1,NEND1
      IF(R(I,1))420,430,440
  420 ISIGN=IMIN
      GO TO  450
  430 ISIGN=IZER
      GO TO  450
  440 ISIGN=IPLUS
  450 NEND=NDIM-INT((FLOAT(I)-1.0)/2.0+1.0E-4)
      IF(I-NROW(K))490,460,490
  460 K=K+1
      KLESS1=KLESS1+1
      IF(NEND-8)470,470,480
  470 WRITE(6,  40)KLESS1,I,ISIGN,(R(I,J),J=1,NEND)
      GO TO  520
  480 WRITE(6,  50)KLESS1,I,ISIGN,(R(I,J),J=1,NEND)
      GO TO  520
  490 IF(NEND-8)500,500,510
  500 WRITE(6,  20)I,ISIGN,(R(I,J),J=1,NEND)
      GO TO  520
  510 WRITE(6,  30)I,ISIGN,(R(I,J),J=1,NEND)
  520 CONTINUE
      FICHNG=SNGL(RTSQ)
      IF(NDIM.GT.NDIMA)WRITE(6,2010)FICHNG
      WRITE(6,  60)
C     DETERMINE NUMBER OF SIGN VARIATIONS IN FIRST COLUMN
  530 ICHNG=NSIGNV(R1,8,NEND1)
      IF(ICHNG-NDIM-2)550,540,540
C     SIGN VARIATIONS SHOULD BE AT LEAST N-1 FOR POSITIVITY
  540 NONNEG=6
      GO TO  630
C
C     SILJAK'S TEST FOR NON-NEGATIVITY
C
  550 DO  560 I=2,IZERO
      J=NROW(I)-NROW(I-1)
      K=NROW(I-1)+1
  560 NV(I)=NSIGNV(R1(K),8,J)
C     TEST ODD SECTOR NUMBERS
```

```
      NPI=0
      DO  570 I=1,IZERO,2
      IF(I+1-IZERO)570,580,590
  570 NPI=2*NROW(I+1)-NROW(I+2)-NROW(I)+2*(NV(I+2)-NV(I+1))+NPI
  580 NPI=NROW(IZERO)-NROW(IZERO-1)-2*NV(IZERO)+NPI
  590 IF(NPI)610,600,610
  600 NONNEG=7
      GO TO  630
  610 NONNEG=4
      GO TO  630
  620 NONNEG=0
  630 IF(IPRT)640,650,640
  640 J=NONNEG+1
      WRITE(6,  80)IOUT(1,J),(IOUT(I,J),I=2,12)
  650 RETURN
      END
```

```
      FUNCTION NSIGNV(X,W,LENGTH)
C     FUNCTION NSIGNV COMPUTES THE NUMBER OF SIGN VARIATIONS
C          IN A VECTOR X, LENGTH LONG
C          W IS A SINGLE PRECISION WORK VECTOR OF LENGTH   LENGTH
C          NSIGNV IS USED IN CONJUNCTION WITH FUNCTION NONNEG
      DIMENSION X(1),W(1)
C     COUNT NUMBER OF SIGN VARIATIONS
      NSIGNV=0
      NDIM=0
      DO   20 I=1,LENGTH
      IF(X(I))10,20,10
   10 NDIM=NDIM+1
      W(NDIM)=SIGN(1.0,X(I))
   20 CONTINUE
      DO   40 I=2,NDIM
      IF(W(I-1)*W(I))30,40,40
   30 NSIGNV=NSIGNV+1
   40 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE SPFCTR(PSI,FACTR,NDIMA,IER)
C     SUBROUTINE SPFCTR COMPUTES THE SPECTRAL FACTOR X(JW) OF
C                                          2
C            AN EVEN POLYNOMIAL Z(W ),I.E.
C
C                           2            2                 *
C                       Z(W ) =   X(JW)     = X(JW)X(JW)
C
C                                          2                         2(N-1)
C     A2K  = INPUT POLYNOMIAL(Z(W ))  A2K(N) = COEF. OF W
C                                                                         N-1
C     FACTR= SPECTRAL FACTOR POLYNOMIAL(X(S))  FACTR(N) = COEF. OF S
C     NDIMA= NUMBER OF COEFICIENTS OF FACTR (ORDER + 1)
C     IER  = ERROR CODE
C            IER = 0  NO ERROR
C            IER = 1  MAX. NO. OF ATTEMPTS AT FINDING QUAD. FACTORS
C                     EXCEEDED
C            IER = 2  ORDER OF INPUT POLYNOMIAL HAS BEEN REDUCED BY
C                    ROUNDOFF OR OVERFLOW OCCURRED DURING NORMALIZATION
C            IER = 3  REMAINING LINEAR FACTOR HAS POSITIVE ZERO
C
C  REMARKSO
C     1. IF FACTORIZATION IS SUCCESSFUL(IER = 0), FACTR MAY STILL NOT
C        BE THE SPECTRAL FACTOR IF THE INPUT POLYNOMIAL IS NOT NON-
C        NEGATIVE DEFINITE  IT WILL DIFFER FROM THE SPECTRAL FACTOR
C        IN THAT THE RIGHT HALF PLANE ZEROS OF THE SPEC. FACTOR WILL
C        BE REFLECTED ABOUT THE IMAGINARY AXIS.  FACTR WILL ALWAYS HAVE
C        LEFT HALF PLANE ZEROS.
C     2. SPFCTR USES SUBROUTINE QFACT TO DETERMINE QUADRATIC FACTORS.
C     3. PSI AND FACTR ARE DOUBLE PRECISION ARRAYS
C                                                 J.M. ELDER   4/23/71
      DOUBLE PRECISION FACTR(1),PSI(1),WORK(20),QUAD(4),
     1A0,A1,R0,R1,F0,F1,DISC,DNORM
      EQUIVALENCE (A0,QUAD(1)),(A1,QUAD(2)),(R0,QUAD(3)),(R1,QUAD(4))
      COMMON WORK,QUAD,F0,F1,DISC,STRT,DEL
      NDIM=NDIMA
      DNORM=1.0D0
      DO   10 I=1,NDIM
   10 FACTR(I)=0.0D0
C     SET NMAX,MAX. NO. OF TRYS AT DETERMINING ALL THE FACTORS OF PSI
C        (PRESENTLY TWICE NDIM) AND ITMAX,MAX.NO. OF ITERATIONS PER TRY
      NMAX=2*NDIM
      NTRY=1
      ITMAX=100
      FACTR(1)=1.0D0
      NDIMF=1
      STRT=1.0E+4
      DEL=STRT/FLOAT(NMAX)
C     GENERATE A NEW GUESS,IF REQUIRED (A1 0,TO TRY TO FORCE REAL
C        DISTINCT ZEROS OF QUADRATIC FACTOR OF COMPANION POLYNOMIAL)
   20 A0=DBLE(STRT)
      A1=2.5D0*DBLE(STRT)
      STRT=STRT-DEL
C     HAS MAXIMUM NUMBER OF ATTEMPTS AT FACTORING PSI BEEN EXCEEDED
   30 IF(NTRY-NMAX)40,40,340
   40 NTRY=NTRY+1
C
C     APPROX. QUAD. FACTOR OF COMPANION POLY. (PSI) USING BAIRSTOW'S M.
      CALL QFACT(PSI,NDIM,QUAD,ITMAX,DNORM,IERQ)
C     TEST FOR ERRORS IN COMPUTING FACTOR
      IF(IERQ)50,60,20
   50 IF(IERQ+2)20,260,350
C     MULTIPLY FACTR BY LHP QUADRATIC PART OF QUARTIC FACTOR
C        COMPUTE LHP QUADRATIC PART COEFFICIENTS
```

```
   60 F1=0.0D0
      JMP=1
C     TEST DISCRIMINANT TO DETERMINE WHERE ROOTS FALL
      DISC=A1**2-4.0D0*A0
      IF(DABS(DISC)-1.0D-10)70,70,80
   70 DISC=0.0D0
   80 IF(DISC)190,170,90
C    PAIR OF REAL ZEROS
   90 IF(A0)100,130,130
C    ZEROS IN OPPOSITE HALF PLANES
C         MULTIPLY IN EXTRA FACTOR
  100 A0=DSQRT(DISC)
      F0=0.5D0*DABS(A1+A0)
      JMP=0
      GO TO  280
C         DIVIDE REDUNDANT FACTOR OUT OF PSI
  110 F0=-0.5D0*(A1-A0)
      R0=PSI(NDIM)
      NDIM=NDIM-1
      J=NDIM
  120 R1=PSI(J)
      PSI(J)=R0
      R0=R1-R0*F0
      J=J-1
      IF(J)200,200,120
C     REAL DISTINCT ZEROS IN SAME HALF PLANE
  130 IF(A1)140,190,190
C     REAL DISTINCT ZEROS IN RIGHT HALF PLANE
  140 IF(JMP-2)150,160,150
  150 A0=DSQRT(DISC)
      F0=-0.5D0*(A1-A0)
      JMP=2
      GO TO  210
  160 F0=-0.5D0*(A1+A0)
      GO TO  200
C     ZERO OF MULTIPLICITY TWO
  170 F0=0.5D0*DABS(A1)
      IF(A1)210,210,180
  180 F1=DSQRT(2.0D0*A1)
      GO TO  200
C     PAIR OF COMPLEX ZEROS OR REAL ZEROS IN LEFT HALF PLANE
  190 F0=DSQRT(A0)
      F1=DSQRT(A1+2.0D0*F0)
C     MULTIPLY IN NEW SPECTRAL FACTOR
  200 JMP=1
  210 NDIMF=NDIMF+2
      FACTR(NDIMF)=1.0D0
      NEND=NDIMF-3
      IF(NEND)240,240,220
  220 DO  230 I=1,NEND
      J=NDIMF-I
  230 FACTR(J)=FACTR(J-2)+F1*FACTR(J-1)+F0*FACTR(J)
  240 FACTR(2)=F0*FACTR(2)+F1*FACTR(1)
      FACTR(1)=F0*FACTR(1)
C     IS SPECTRAL FACTOR COMPLETE
      IF(NDIM-1)250,320,250
    2 GO TO (20,30),JMP
C     MULTIPLY IN FIRST ORDER FACTOR
  260 F0=PSI(1)
  270 IF(F0)360,280,280
  280 F0=DSQRT(F0)
      NDIMF=NDIMF+1
      FACTR(NDIMF)=1.0D0
      NEND=NDIMF-2
      IF(NEND)310,310,290
  290 DO  300 I=1,NEND
```

```
      J=NDIMF-I
  300 FACTR(J)=F0*FACTR(J)+FACTR(J-1)
  310 FACTR(1)=F0*FACTR(1)
      IF(JMP)320,110,320
  320 IER=0
C     UNNORMALIZE FACTR AND PSI
      DISC=DSQRT(DNORM)
      DO  330 I=1,NDIMA
      FACTR(I)=DISC*FACTR(I)
  330 PSI(I)=DNORM*PSI(I)
      RETURN
C     MAXIMUM NUMBER OF ATTEMPTS AT FACTORING PSI EXCEEDED
  340 IER=1
      RETURN
C     ORDER OF LAST FACTOR OF PSI REDUCED BY ROUNDOFF OR OVERFLOW
C        OCCURRED DURING NORMALIZATION OF POLYNOMIAL
  350 IER=2
      RETURN
C     REMAINING LINEAR FACTOR OF COMPANION POLY. HAS POSITIVE ZERO
  360 IER=3
      RETURN
      END
```

```
      SUBROUTINE QFACT(C,IC,Q,LIM,DNORM,IER)
C     SUBROUTINE QFACT COMPUTES AND REMOVES QUADRATIC FACTORS FROM
C        REAL POLYNOMIALS USING BAIRSTOW'S METHOD.  IT IS A MODIFICATION
C        OF SUBROUTINE DPQFB OF THE SYSTEMS 360 SCIENTIFIC SUBROUTINE
C        PACKAGE.
C           C    - DOUBLE PRECISION INPUT VECTOR CONTAINING THE
C                  COEFFICIENTS OF P(X) - C(1) IS THE CONSTANT TERM
C                  (DIMENSION IC)
C           IC   - DIMENSION OF C
C           Q    - DOUBLE PRECISION VECTOR OF DIMENSION 4 - ON INPUT Q(1)
C                  AND Q(2) CONTAIN INITIAL GUESSES FOR Q1 AND Q2 - ON
C                  RETURN Q(1) AND Q(2) CONTAIN THE REFINED COEFFICIENTS
C                  Q1 AND Q2 OF Q(X), WHILE Q(3) AND Q(4) CONTAIN THE
C                  COEFFICIENTS A AND B OF A+B*X, WHICH IS THE REMAINDER
C                  OF THE QUOTIENT OF P(X) BY Q(X)
C           LIM  - INPUT VALUE SPECIFYING THE MAXIMUM NUMBER OF
C                  ITERATIONS TO BE PERFORMED
C           DNORM-NONZERO COEFFICIENT OF HIGHEST POWER TERM,BY WHICH
C                  C IS NORMALIZED
C           IER  - RESULTING ERROR PARAMETER (SEE REMARKS)
C                  IER= 0 - NO ERROR
C                  IER= 1 - NO CONVERGENCE WITHIN LIM ITERATIONS
C                  IER=-1 - THE POLYNOMIAL P(X) IS CONSTANT OR UNDEFINED
C                         - OR OVERFLOW OCCURRED IN NORMALIZING P(X)
C                  IER=-2 - THE POLYNOMIAL P(X) IS OF DEGREE 1
C                  IER=-3 - NO FURTHER REFINEMENT OF THE APPROXIMATION TO
C                           A QUADRATIC FACTOR IS FEASIBLE, DUE TO EITHER
C                           DIVISION BY 0, OVERFLOW OR AN INITIAL GUESS
C                           THAT IS NOT SUFFICIENTLY CLOSE TO A FACTOR OF
C                           P(X)
C                                           J.M. ELDER 4/29/71
      DIMENSION C(1),Q(1)
      DOUBLE PRECISION A,B,AA,BB,CA,CB,CC,CD,A1,B1,C1,H,HH,Q1,Q2,QQ1,
     1                 QQ2,QQQ1,QQQ2,DQ1,DQ2,EPS,EPS1,C,Q,DNORM
C
C        TEST ON LEADING ZERO COEFFICIENTS
      IER=0
      J=IC+1
   10 J=J-1
      IF(J-1)420,420,20
   20 IF(C(J))30,10,30
C
C        NORMALIZATION OF REMAINING COEFFICIENTS
   30 A=C(J)
      IF(A-1.D0)40,60,40
   40 DO   50 I=1,J
      C(I)=C(I)/A
      CALL OVERFL(N)
      IF(N-2)420,50,50
   50 CONTINUE
      DNORM=A
C
C        TEST ON NECESSITY OF BAIRSTOW ITERATION
   60 IF(J-3)430,380,70
C
C        PREPARE BAIRSTOW ITERATION
   70 EPS=1.D-14
      EPS1=1.D-6
      L=0
      LL=0
      Q1=Q(1)
      Q2=Q(2)
      QQ1=0.D0
      QQ2=0.D0
```

```
      AA=C(1)
      BB=C(2)
      CB=DABS(AA)
      CA=DABS(BB)
      IF(CB-CA)80,90,100
   80 CC=CB+CB
      CB=CB/CA
      CA=1.D0
      GO TO  110
   90 CC=CA+CA
      CA=1.D0
      CB=1.D0
      GO TO  110
  100 CC=CA+CA
      CA=CA/CB
      CB=1.D0
  110 CD=CC*.1D0
C
C        START BAIRSTOW ITERATION
C        PREPARE NESTED MULTIPLICATION
  120 A=0.D0
      B=A
      A1=A
      B1=A
      I=J
      QQQ1=Q1
      QQQ2=Q2
      DQ1=HH
      DQ2=H
C
C        START NESTED MULTIPLICATION
  130 H=-Q1*B-Q2*A+C(I)
      CALL OVERFL(N)
      IF(N-2)440,140,140
  140 B=A
      A=H
      I=I-1
      IF(I-1)180,150,160
  150 H=0.D0
  160 H=-Q1*B1-Q2*A1+H
      CALL OVERFL(N)
      IF(N-2)440,170,170
  170 C1=B1
      B1=A1
      A1=H
      GO TO  130
C        END OF NESTED MULTIPLICATION
C
C        TEST ON SATISFACTORY ACCURACY
  180 H=CA*DABS(A)+CB*DABS(B)
      IF(LL)190,190,390
  190 L=L+1
      IF(DABS(A)-EPS*DABS(C(1)))200,200,210
  200 IF(DABS(B)-EPS*DABS(C(2)))390,390,210
C
C        TEST ON LINEAR REMAINDER OF MINIMUM NORM
  210 IF(H-CC)220,220,230
  220 AA=A
      BB=B
      CC=H
      QQ1=Q1
      QQ2=Q2
C
C        TEST ON LAST ITERATION STEP
  230 IF(L-LIM)280,280,240
C
```

```
C         TEST ON RESTART OF BAIRSTOW ITERATION WITH ZERO INITIAL GUESS
  240 IF(H-CD)450,450,250
  250 IF(Q(1))270,260,270
  260 IF(Q(2))270,440,270
  270 Q(1)=0.D0
      Q(2)=0.D0
      GO TO   70
C
C         PERFORM ITERATION STEP
  280 HH=DMAX1(DABS(A1),DABS(B1),DABS(C1))
      IF(HH)440,440,290
  290 A1=A1/HH
      B1=B1/HH
      C1=C1/HH
      H=A1*C1-B1*B1
      IF(H)300,440,300
  300 A=A/HH
      B=B/HH
      HH=(B*A1-A*B1)/H
      H=(A*C1-B*B1)/H
      Q1=Q1+HH
      Q2=Q2+H
C         END OF ITERATION STEP
C
C         TEST ON SATISFACTORY RELATIVE ERROR OF ITERATED VALUES
      IF(DABS(HH)-EPS*DABS(Q1))310,310,330
  310 IF(DABS(H -EPS*DABS(Q2))320,320,330
  320 LL=1
      GO TO  120
C
C         TEST ON DECREASING RELATIVE ERRORS
  330 IF(L-1)120,120,340
  340 IF(DABS(HH)-EPS1*DABS(Q1))350,350,120
  350 IF(DABS(H)-EPS1*DABS(Q2))360,360,120
  360 IF(DABS(QQQ1*HH)-DABS(Q1*DQ1))370,460,460
  370 IF(DABS(QQQ2*H)-DABS(Q2*DQ2))120,460,460
C         END OF BAIRSTOW ITERATION
C
C         EXIT IN CASE OF QUADRATIC POLYNOMIAL
  380 Q(1)=C(1)
      Q(2)=C(2)
      Q(3)=0.D0
      Q(4)=0.D0
      IC=J-2
      RETURN
C
C         EXIT IN CASE OF SUFFICIENT ACCURACY
  390 Q(1)=Q1
      Q(2)=Q2
      Q(3)=A
      Q(4)=B
C     DIVIDE OUT QUADRATIC FACTOR
      A1=C(IC)
      B1=C(IC-1)
      J=IC-2
      IC=IC-2
  400 C1=A1
      A1=B1-C1*Q2
      B1=C(J)-C1*Q1
      C(J)=C1
      J=J-1
      IF(J-1)400,410,400
  410 C(J)=A1
      RETURN
C
C         ERROR EXIT IN CASE OF ZERO OR CONSTANT POLYNOMIAL
```

```
  420 IER=-1
      RETURN
C
C           ERROR EXIT IN CASE OF LINEAR POLYNOMIAL
  430 IER=-2
      RETURN
C
C           ERROR EXIT IN CASE OF NONREFINED QUADRATIC FACTOR
  440 IER=-3
      GO TO  460
C
C           ERROR EXIT IN CASE OF UNSATISFACTORY ACCURACY
  450 IER=1
  460 Q(1)=QQ1
      Q(2)=QQ2
      IC=J-1
      Q(3)=AA
      Q(4)=BB
      RETURN
      END


      SUBROUTINE MAGSQ(X,Z,N)
C     SUBROUTINE MAGSQ COMPUTES THE MAGNITUDE SQUARE POLYNOMIAL,I.E.
C
C                               2             2            *
C                         Z(W  )  =  X(JW)     = X(JW)X(JW)
C
C     X = INPUT POLYNOMIAL
C                                                       2(N-1)
C     Z = MAGNITUDE SQUARE POLYNOMIAL  Z(N) = COEF. OF W
C     N = NUMBER OF COEFICIENTS (ORDER + 1)
C                                                J.M. ELDER   4/20/71
      DOUBLE PRECISION X,Z
      DIMENSION X(1),Z(1)
      DO    10 I=1,N
   10 Z(I)=0.0
      DO    50 I=1,N
      DO    50 J=1,N
      K=I+J
C     TEST FOR ODD K (TERM OF Z+2)...KEEP ONLY EVEN K
      FK=FLOAT(K)/2.0+1.0E-4
      IF(K-2*INT(FK))50,20,50
   20 K=INT(FK)
C     DETERMINE SIGN OF SUM TERM...CHANGE SIGN IF (I+3J)/2 = K+J ODD
      FK=FLOAT(K+J)/2.0+1.0E-4
      IF(2*INT(FK)-K-J)30,40,30
   30 Z(K)=-X(I)*X(J)+Z(K)
      GO TO    50
   40 Z(K)=X(I)*X(J)+Z(K)
   50 CONTINUE
      RETURN
      END
```

```
      DOUBLE PRECISION PHI,PHIK,PSI,FACTR,VAL,D1,D2
      DIMENSION PHI(20),PHIK(20),PSI(20),FACTR(20),A2K(20)
      COMMON A2K,PHI,PHIK,D1,D2
      DATA IBLK/3H   /,IP/3HP  /,IPK/3HPK /,IEND/3HEND/,IGO/3H*GO/
   10 FORMAT(1A3,I2,D25.16)
   20 FORMAT(//11X,26HCHARACTERISTIC POLYNOMIALS/39X,1HN/22X,
     117HCOEFFICIENTS OF S/6X,1HN,5X,14HOPEN LOOP(PHI),4X,
     217HCLOSED LOOP(PHIK))
   30 FORMAT(I7,1P2D19.7)
   40 FORMAT(//5X,10HDIAGONAL Q/5X,17HDIAGONAL TERMS.../(22X,1PD19.12))
   50 FORMAT(//5X,11HUNIT RANK Q/5X,25HAVERAGE ERROR IN FACTOR =,
     11PD11.3/20X,6HFACTOR,20X,3HPSI,21X,5HCHECK/(10X,1P3D25.12))
   60 FORMAT(5X,50H***** POLYNOMIAL HAS HIGHEST POWER TERM ZERO *****)
   70 FORMAT(5X,46H***** INPUT CONTAINS UNIDENTIFIBLE LABEL *****)
   80 FORMAT(5X,47H***** SPECTRAL FACTORIZATION UNSUCCESSFUL *****)
   90 FORMAT(5X,35H***** SYSTEM PAIR NON-OPTIMAL *****)
  100 FORMAT(1H ,A3,I2,D25.16)
  110 FORMAT(4H PHI)
  120 FORMAT(5H PHIK)
      NDIM=4
      PHI(1)=5.0
      PHI(2)=9.0
      PHI(3)=5.0
      PHI(4)=1.0
      PHIK(1)=26.0
      PHIK(2)=25.0
      PHIK(3)=8.0
      PHIK(4)=1.0
      GO TO  280
  130 DO  140 I=1,20
      PHI(I)=0.000
  140 PHIK(I)=0.000
  150 JMP=-1
      NDIM=0
      DECODE INPUT
  160 READ(5,  10)ICHAR,I,VAL
      WRITE(6, 100)ICHAR,I,VAL
      IF(ICHAR-IBLK)170,210,170
  170 IF(ICHAR-IP)180,240,180
  180 IF(ICHAR-IPK)190,220,190
  190 IF(ICHAR-IGO)200,280,200
  200 IF(ICHAR-IEND)420,430,420
  210 IF(JMP)160,250,230
  220 JMP=1
  230 PHIK(I)=VAL
      WRITE(6, 120)
      GO TO  260
  240 JMP=0
  250 PHI(I)=VAL
      WRITE(6, 110)
  260 IF(NDIM-I)270,160,160
  270 NDIM=I
      GO TO  160
      NORMALIZE POLYNOMIALS SO THAT HIGHEST POWER TERM IS ONE
  280 NORD=NDIM-1
      D1=PHI(NDIM)
      D2=PHIK(NDIM)
      IF(DABS(D1)-1.0D-40)410,410,290
  290 IF(DABS(D2)-1.0D-40)410,410,300
  300 DO  310 I=1,NORD
      PHI(I)=PHI(I)/D1
  310 PHIK(I)=PHIK(I)/D2
      PHI(NDIM)=1.000
      PHIK(NDIM)=1.000
```

```
      WRITE(6,  20)
      DO  320 I=1,NDIM
      J=I-1
  320 WRITE(6,  30)J,PHI(I),PHIK(I)
C                                                      2               2
C     COMPUTE MAGNITUDE SQUARE POLYNOMIALS  PSI= PHI      PHI= PHIK
  330 CALL MAGSQ(PHI,PSI,NDIM)
      CALL MAGSQ(PHIK,PHI,NDIM)
C
C     COMPUTE DIFFERENCES OF MAGNITUDE SQUARE POLYNOMIALS
      DO  340 I=1,NDIM
      PSI(I)=PHI(I)-PSI(I)
  340 A2K(I)=SNGL(PSI(I))
C                                                      2
C     IS THE CLOSED-LOOP SYSTEM OPTIMAL,I.E. IS PSI(W ) POSITIVE SEMID
      ITEST=NONNEG(A2K,NORD,1)
      IF(ITEST-5)390,350,360
C     DIAGONAL Q
  350 WRITE(6,  40)(PSI(I),I=1,NORD)
C     RANK ONE Q
  360 CALL SPFCTR(PSI,FACTR,NORD,IER)
      IF(IER)400,370,400
C     CHECK FACTORIZATION
  370 CALL MAGSQ(FACTR,PHI,NORD)
      D1=0.0D0
      DO  380 I=1,NORD
  380 D1=D1+DABS(PSI(I)-PHI(I))
      D1=D1/DBLE(FLOAT(NORD))
      WRITE(6,  50)D1,(FACTR(I),PSI(I),PHI(I),I=1,NORD)
      GO TO  130
  390 WRITE(6,  90)
      GO TO  130
  400 WRITE(6,  80)
      GO TO  130
  410 WRITE(6,  60)
      GO TO  130
  420 WRITE(6,  70)
      GO TO  130
  430 STOP
      END
```

PROGRAM OPTI

CHARACTERISTIC POLYNOMIALS

N	COEFFICIENTS OF S^N OPEN LOOP(PHI)	CLOSED LOOP(PH K)
0	2.4960000D 04	2.4960000D 04
1	4.4128000D 04	4.4221787D 04
2	3.2304000D 04	3.2469980D 04
3	1.2796000D 04	1.2916180D 04
4	2.9820000D 03	3.0276270D 03
5	4.1000000D 02	4.1937845D 02
6	3.1000000D 01	3.1870941D 01
7	1.0000000D 00	1.0000000D 00

NORMALIZED ROUTH ARRAY

	ROW#		NORMALIZED ROWS						
*R 1	1	+	4.29E-01	1.00E 00	5.83E-01	-1.43E-01	-1.67E-01	0.0	1.19E-02
	2	+	5.14E-01	1.00E 00	4.67E-01	-8.57E-02	-6.67E-02	0.0	0.0
	3	+	8.57E-01	1.00E 00	-3.67E-01	-5.71E-01	0.0	6.12E-02	
	4	+	5.82E-01	1.00E 00	3.74E-01	-9.70E-02	-5.35E-02	0.0	
	5	-	-5.14E-01	-1.00E 00	-4.07E-01	8.57E-02	6.67E-02		
	6	-	-8.57E-01	-1.00E 00	-1.64E-16	1.43E-01	0.0		
*R 2	7	-	-8.57E-01	-1.00E 00	-2.43E-16	1.43E-01			
	8	-	-1.00E 00	-7.78E-01	-9.44E-17	0.0			
	9	-	-1.00E 00	-4.86E-16	4.29E-01				
	10	-	-1.00E 00	-5.51E-01	0.0				
	11	+	1.00E 00	7.78E-01					
	12	+	1.00E 00	0.0					
*R 3	13	+	1.00E 00						

* ROWS PRECEEDING ZERO ROWS
(ZERO ROWS HAVE BEEN RESOLVED BY DIFFERENTIATING PRECEEDING ROW)

PSI(W^2) IS...... POSITIVE SEMI DEFINITE, POLYNOMIAL PASSES SUFFICIENCY TEST

UNIT RANK Q
AVERAGE ERROR IN FACTOR = 6.122D-15

FACTOR	PSI	CHECK
1.000000000000D 00	1.000000000000D 00	1.000000000000D 00
3.464101615138D 00	0.0	4.440892098501D-15
6.000000000000D 00	-1.400000000000D 01	-1.400000000000D 01
1.039230484541D 01	1.200000000000D 01	1.200000000000D 01
1.100000000000D 01	4.900000000000D 01	4.900000000000D 01
6.928203230276D 00	-8.400000000000D 01	-8.400000000000D 01
6.000000000000D 00	3.600000000000D 01	3.600000000000D 01

REFERENCES

A1 B. D. O. Anderson and J. B. Moore, Linear Optimal Control, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971.

A2 F. Ayers, Theory and Problems of Matrices, Schaum Publishing Co., New York, N.Y., 1962.

A3 A. A. Albert, Fundamental Concepts of Higher Algebra, University of Chicago Press, Chicago, Ill., 1956.

A4 B. D. O. Anderson and J. B. Moore, "Linear System Optimization with Prescribed Degree of Stability," Proc. of the IEE(GB), Vol. 116, No. 2, pp. 2083-2087, Dec. 1969.

B1 A. E. Bryson and Y. C. Ho, Applied Optimal Control, Blaisdell Publishing Co., Waltham, Mass., 1969.

B2 R. W. Brockett, Finite Dimensional Linear Systems, John Wiley and Sons, Inc., New York, N.Y., 1970.

B3 J. V. Breakwell and Y. C. Ho, "On the Conjugate Point Condition for the Control Problem," Int. J. Engng. Sci., Vol. 2, pp. 565-579, 1965.

B4 T. E. Bullock, "Computation of Optimal Controls by a Method Based on Second Variations," Tech. Rep. to N.A.S.A., Grant No. NsG 133-61 (SUDAAR No. 297), Stanford University, Stanford, Cal., Dec. 1966.

B5 R. Bellman, Dynamic Programming, Princeton Univ. Press, Princeton, N.J., 1957.

B6 R. S. Bucy, private correspondence, Dept. of Aerospace Engineering, Univ. of S. Cal., Los Angeles, Cal.

B7 T. E. Bullock and C. E. Fosha, "LQL - A General Purpose Fortran Program for Control Estimation and Simulation," Conf. Rec. Fourth Asilomar Conf. on Circuits and Systems," Asilomar, Cal, Nov. 1970.

B8 H. W. Bode, Network Analysis and Feedback Amplifier Design, Van Nostrand and Co., Inc., New York, N.Y., 1945.

B9 W. H. Boykin, T. E. Bullock, J. M. Elder, B. D. Frazier, "Digital Missile Controller Design and Analysis Information," Tech. Rep. to U.S. Army Missile Cmmd., Redstone Arsenal, Ala., Contr. No. DAAH01-70-C-0639, University of Florida, Gainesvilla, Fla., May 1971.

C1 S. S. L. Chang, Synthesis of Optimum Control Systems, McGraw-Hill Book Co., New York, N.Y., 1961.

C2 C. G. Cullen, Matrices and Linear Transformations, Addison-Wesley Publishing Co., Reading, Mass., 1966.

C3 S. S. L. Chang, "Root-Square-Locus Plot - A Geometric Method for the Synthesis of Optimal Servo Systems," IRE Conv. Record, 1958.

D1 H. T. Davis, Introduction to Nonlinear Differential and Integral Equations, U. S. Atomic Energy Commission, Washington, D.C., 1960.

D2 H. D'Angelo, Linear Time-Varying Systems: Analysis and Synthesis, Allyn and Bacon Publishing Co., Boston, Mass., 1969.

D3 P. M. De Russo, R. J. Roy, C. M. Close, State Variables for Engineers, John Wiley and Sons, Inc., New York, N.Y., 1965.

D4 R. C. Dorf, Modern Control Systems, Addison-Wesley Publishing Co., Reading, Mass., 1967.

E1 J. M. Elder and T. E. Bullock, "Quadratic Performance Index Generation for Optimal Regulator Design," Proceedings, 1971 IEEE Decision and Control Conference, Miami Beach, Fla., Dec. 15-17, 1971.

E2 O. I. Elgerd, Control Systems Theory, McGraw-Hill Book Co., New York, N.Y., 1967.

E3 J. M. Elder and T. E. Bullock, "On the Inverse Linear Plant Quadratic Loss Optimization Problem," Proceedings, Third Southeastern Symposium on System Theory, Atlanta, Ga., April 5-6, 1971.

G1 I. M. Gelfand and S. V. Fomin, Calculus of Variations, translated by R. A. Silverman, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

G2 D. Graham and R. C. Lanthrop, "The Synthesis of Optimum Transient Response: Criteria and Standard Forms," Trans. AIEE, Vol. 72, pt. 2, pp. 278-288, Nov. 1953.

G3 A. Gill and R. Sivan, "Optimal Control of Linear Systems with Quadratic Costs Which Are Not Necessarily Positive Definite," IEEE Trans. on Auto. Control, Vol. AC-14, No. 1, pp. 83-85, Feb. 1969.

G4 F. R. Gantmacher, The Theory of Matrices, Chelsea Publishing Co., New York, N.Y., 1959.

G5 S. Gerschgorin, "Uber die Abgrenzung der Eigenwerte einer Matrix," Izv. Akad. Nauk SSSR, Ser. Fiz.-Mat., Vol. 6, pp. 749-754, 1931.

G6 E. A. Guillemin, Synthesis of Passive Networks, John Wiley and Sons, Inc., New York, N.Y., 1957.

H1 F. E. Hohn, Elementary Matrix Algebra, The Macmillan Co., New York, N.Y., 1958.

H2 C. H. Houpis and C. T. Constantinides, "Correlation between Conventional Control Figures of Merit and the Q Matrix of the Quadratic Cost Function: Third-Order Plant," Proceedings of the 14th Midwest Symposium on Circuit Theory, May, 1971.

H3 F. B. Hildebrand, Introduction to Numerical Analysis, McGraw-Hill Book Co., New York, N.Y., 1956.

I1 IBM Corp., System/360 Scientific Subroutine Package (360A-CM-03X) Version III Programmer's Manual, IBM Corp., Data Processing Division, White Plains, N.Y.

J1 C. D. Johnson and W. M. Wonham, "Another Note on the Transformation to Canonical (Phase-Variable) Form," IEEE Trans. on Auto. Control, Vol. AC-11, No. 3, pp. 609-610, July 1966.

J2 E. I. Jury, "A Note on the Absolute Stability Test," Proceedings of the IEEE, Vol. 58, No. 5, May 1970, pp. 823 ff.

K1 R. E. Kalman, "When Is a Linear Control System Optimal?", Trans. ASME, Ser. D, J. of Basic Engng., Vol. 86, pp. 51-60, March 1964.

K2 R. E. Kalman, "Contributions to the Theory of Optimal Control," Bol. Soc. Mat. Mexicana, Vol. 5, pp. 102-119, 1960.

K3 R. E. Kalman and T. S. Englar, "A User's Manual for ASPC," Tech. Rep. to NASA, Ames Research Center, Moffett Field, Cal., Contr. No. NAS2-1107, RIAS Div. of Martin Marietta Corp., Baltimore, Md., NASA CR-475, June 1966.

K4 E. Kreindler and J. K. Hedrick, "On Equivalence of Quadratic Loss Functions," Int. J. of Control, Vol. 11, No. 2, pp. 213-222, 1970.

K5 R. E. Kalman, "Mathematical Description of Linear Dynamical Systems," SIAM J. Control, Ser. A, Vol. 1, No. 2, pp. 152-192, 1963.

K6 R. E. Kalman, class notes EE 692, University of Florida, Gainesville, Fla., Fall 1971.

K7 B. C. Kuo, Automatic Control Systems, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1962.

K8 R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, Ser. D, J. of Basic Engng., Vol. 82, pp. 34-45, March 1960.

K9 R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," Trans. ASME, Ser. D, J. of Basic Engng., Vol. 83, pp. 95-108, March 1961.

K10 J. S. Karmarkar, "On Nonnegativity of Even Polynomials," Proc. of the IEEE, Vol. 58, No. 5, pp. 835 ff, May 1970.

K11 L. G. Kelly, Handbook of Numerical Methods and Applications, Addison-Wesley Publishing Co., Reading, Mass., 1967.

K12 B. C. Kuo, Analysis and Synthesis of Sampled-Data Control Systems, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

L1 R. J. Leake, "Return Difference Bode Diagram for Optimal System Design," IEEE Trans. on Auto. Control, Vol. AC-10, No. 3, pp. 342-344, July 1965.

L2 I. Lee, "Optimal Trajectory, Guidance, and Conjugate Points," Information and Control, Vol. 8, pp. 589-606, 1965.

L3 A. A. Lyapunov, Probléme Général de la Stabilité du Movement, Ph.D. Thesis, 1892, Annuals of Math. Study No. 17, Princeton Univ. Press, Princeton, N.J., 1949.

L4 J. P. LaSalle and S. Lepschetz, Stability by Lyapunov's Direct Method, Academic Press, New York, N.Y., 1961.

M1 C. C. Macduffee, The Theory of Matrices, Chelsea Publishing Co., New York, N.Y., 1946.

N1 G. C. Newton, L. A. Gould, J. F. Kaiser, Analytical Design of Linear Feedback Controls, John Wiley and Sons, Inc., New York, N.Y., 1957.

P1 J. B. Pearson, "Compensator Design for Dynamic Optimization," Int. J. of Control, Vol. 9, No. 4, pp. 473-482, 1969.

P2 J. B. Pearson and C. Y. Ding, "Compensator Design for Multivariable Linear Systems," IEEE Trans. on Auto. Control, Vol. AC-14, No. 2, pp. 130-134, April 1969.

R1 W. J. Rugh, "On the Characteristics of Linear Optimal Systems," J. Franklin Institute, Vol. 291, No. 5, pp. 371-385, May 1971.

R2 D. S. Rane, "A Simplified Transformation to (Phase-Variable) Canonical Form," IEEE Trans. on Auto. Control, Vol. AC-11, No. 3, pp. 608, July 1966.

R3 B. Ramaswami and K. Ramar, "Transformation to Phase Variable Canonical Form," IEEE Trans. on Auto. Control, Vol. AC-13, No. 6, pp. 746 ff, Dec. 1968.

S1 R. Sivan and S. Shalvi, "On the Existence of Optimal Solutions for the Linear System Quadratic Cost Problem," IEEE Trans. on Auto. Control, Vol. AC-13, No. 2, pp. 188-191, April 1968.

S2 D. G. Schultz and J. L. Melsa, State Functions and Linear Control Systems," McGraw-Hill Book Co., New York, N.Y., 1967.

S3 D. Šiljak, "Nonnegative Polynomials: A Criterion," Proc. of the IEEE, Vol. 58, No. 9, pp. 1370 ff, Sept. 1970.

T1 J. S. Tyler and F. B. Tuteur, "The Use of a Quadratic Performance Index to Design Multivariable Control Systems," IEEE Trans. on Auto. Control, Vol. AC-11, No. 1, pp. 84-92, Jan. 1966.

T2 J. G. Truxal, Control Engineer's Handbook, McGraw-Hill Book Co., New York, N.Y., 1958.

T3 D. F. Tuttle, Network Synthesis, Vol. I, John Wiley and Sons, Inc., New York, N.Y., 1958.

V1 M. E. Van Valkenburg, Introduction to Modern Network Synthesis, John Wiley and Sons, Inc., New York, N.Y., 1960.

W1 N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series, John Wiley and Sons, Inc., New York, N.Y., 1949.

W2 W. M. Wonham and C. D. Johnson, "Optimal Bang-Bang Control with Quadratic Performance Index," Trans. ASME, Ser. D., J. of Basic Engng., Vol. 86, pp. 107-115, March 1964.

Z1 L. A. Zadeh and E. Polak, System Theory, McGraw-Hill Book Co., New York, N.Y., 1969.

BIOGRAPHICAL SKETCH

Joseph Marcus Elder, Jr., the eldest son of Joseph and Marjorie Elder, was born in Richmond, Kentucky, on April 16, 1946. Upon graduation from high school in Cocoa, Florida, in June, 1964, he began university studies at the Georgia Institute of Technology in Atlanta, Georgia.

While enrolled at Georgia Tech he was employed by the Florida Power and Light Company at Cocoa, Florida, in the summer of 1965 and by the IBM Corporation, Federal Systems Division at Kennedy Space Center, Florida, in the summers of 1965 and 1966. In June, 1968, he received the Bachelor of Electrical Engineering degree and began evening graduate study while in the employ of the Lockheed-Georgia Company in Marietta, Georgia. He married the former Miss Dina Phillips of Talladega, Alabama, in August, 1968.

In January, 1969, he became a Graduate Teaching Assistant at Georgia Tech, and in August was awarded a Master of Science in Electrical Engineering degree.

He and his wife, Dina, entered the Graduate School of the University of Florida in September of 1969, where she received the degree of Master of Education with a major in Mathematics in December, 1970, and he is currently completing requirements for the degree of Doctor of Philosophy. While at the University of Florida, Mr. Elder was a Graduate Research Assistant and an NDEA Fellow.

Joseph Marcus Elder, Jr. is a member of the Institute of Electrical and Electronic Engineers and Simulation Councils.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Thomas E. Bullock

Thomas E. Bullock, Chairman
Associate Professor of
Electrical Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William H. Boykin

William H. Boykin
Assistant Professor of
Electrical Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Allen E. Durling

Allen E. Durling
Professor of Electrical Engineering

This dissertation was submitted to the Dean of the College of Engineering and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

March 1972

Dean, College of Engineering

Dean, Graduate School